Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 06

Lecture 28: Hurst Exponent - Estimation under ARFIMA

Hello all, welcome to this course on time series modeling and forecasting using R. Again, this week, the focus has been on designing or discussing a new kind of time series process called the ARFIMA process. Now, again, just to quickly review a few aspects of the ARFIMA process, which we saw in the last couple of lectures. So, firstly, the full form of ARFIMA is nothing but auto-regressive fractionally integrated moving average. Now, in the last lecture, the idea of fractionally integrated should be clear to all of you, I hope.

So, what do you mean by fractionally integrated? Fractionally integrated means that the value D can take need not be an integer, as opposed to what happens in the ARIMA process, right? So, here we talk about special cases. So, D can actually take values between minus half and half, right? Or even bigger than or equal to, let us say, half, right?

But still, the values D can take need not be integers; they can be fractions also. Right. So, what would happen if a time series process is integrated but need not be integrated with an integer order? It may be integrated with a fractional order also. So, let us say 0.5, minus 0.5, or let us say 0.25, minus 0.25, etc.

Right. And especially in the last lecture, we talked about some special cases, right? So, we started this session by defining our FEMA 0D0 process. So, our FEMA 0D0 is nothing but fractionally integrated noise, right? And then we talked about some special cases, such as what would happen if my D is, let us say, negative or what would happen if my D is positive, right?

And towards the end of the last lecture, we talked about a general ARFIMA process where you also have an AR ordering and an MA ordering. And on top of that, the value that D can take can also be a fraction. Now, in today's lecture, we will focus more on a particular kind of exponent called the Hurst exponent, right? And the notation of this

Hurst exponent is capital H, right? And the entire session today will be sort of focusing or revolving around the Hurst exponent.

So, what exactly is the Hurst exponent? So, it is an index for capturing long-range dependency. So, again, as we discussed towards the end of the last lecture, just by looking at the value of this H, or the value of this capital H, or the value of this Hurst exponent, I can gauge the time series' persistence or anti-persistence nature. So, for particular values of H, my time series would become, or it would sort of indicate, let us say, long-term behavior or persistence, and for other values of H, it would indicate an anti-persistent nature. So, a couple of use cases as to where the Hurst exponent might be used.

So, it is used as a measure of the long-term memory of any underlying time series. So, the whole idea is to capture the long-term memory or persistence of any time series process using one fixed value, which is called the Hurst exponent. So, this was developed in hydrology for determining the optimum dam size for the Nile River's rain and drought conditions, right? So, the first practical use case came from a hydrological perspective, where the experimenters wanted to determine the optimum dam size. So, how high should they build the dam or something like that on the banks of the River Nile to monitor the rain conditions and the drought conditions, etc.?

So, just to balance out between, let us say, excessive rains or excessive droughts in that area, they developed the idea of the Hurst exponent. But as opposed to that, this is not just restricted to hydrology, of course. So, the Hurst exponent has been used in fields like finance, hydrology, and physics to understand the behavior of processes over time. So, this is a small historical idea about when the Hurst exponent was developed or, rather, for what it was developed as the very first use case, and then some other use cases where, these days or recently, people have been experimenting with the Hurst exponent. Alright, so firstly, we will try to construct the Hurst exponent.

So, this would involve some notation and some mathematics, and then later on, we will see one particular use case of that also. So, as always, let us say y_t, t equals 1, 2, 3, 4 up to n, be a realization of a time series. So, y_t is nothing but the practical data that is being collected, and let us say the mean of y_t happens to be y_n bar, and then the variance of y_t happens to be s_n squared. So, let us say you collect some practical data, be it coming from any field, let us say finance, hydrology, whatever, but notation-wise, y_t is a realization of a time series with mean y_n bar and variance s_n squared. Then the first ingredient that we require is called Z_t.

And what is Zt or capital Zt? So, capital Zt is nothing but the mean adjusted partial sums. So, the mean adjusted partial sum. So, now again if you notice here, we are putting a subscript on Z also. So, Zt.

$$Z_t = \sum_{j=1}^{t} y_j - t\bar{y}_t\,,\ \ t = 1,2,\dots,n$$

So, subscript is T. So, it sort of depends on the time point. So, Z1, Z2, Z3, etc. up to Zn. But for any subscript, the value that it takes is nothing but a partial sum. So, summation j going from 1 to t yj minus t into the corresponding mean. So, yt bar.

So, just to give you a few initial values of Zt. So, what would be let us say Z1. So, Z1 would be nothing but summation j going from 1 to 1. So, this would be nothing but y1 minus t and then again in this case t is 1. So, y1 minus y1 bar.

Similarly, what would be my z2? So, z2 would be nothing but, let us say, summation. Now, j going from 1 to 2, yj minus now the value of t is 2. So, t y2 or rather 2 into y2 bar, etc. So, for a particular value of t, I can write down those partial sums.

So, these are partial sums because So, in this case, you are simply adding the first sum or the first value and the second value. Similarly, when we create Z3, so Z3 will involve three terms in the partial sum, etc. So, you are sort of building upon one at a time as T increases. And whenever T becomes n, so Zn would be nothing but summation j going from 1 to n and then yj minus n into yn bar.

So, this is the notation for denoting the partial sums here. Now, the second ingredient we will require is called the adjusted range. So, adjusted range is given by this capital Rn, and then the adjusted range definition is quite simple. So, it is nothing but the maximum value of all the values from z1, z2 up to zn minus the minimum value of that same set, right? So, once you find out all these values, right, z1, z2, zn, then the only thing to do is to find out the maximum value of that and the minimum value and then take a difference.

$$R_n = \max\{Z_1, Z_2, \dots, Z_n\} - \min\{Z_1, Z_2, \dots, Z_n\}$$

So, such a difference is called the adjusted range. And in short, or in other words, the rescaled adjusted range. So, what is that? The rescaled adjusted range is nothing but Rn divided by the corresponding standard deviation, or Sn. So, why rescale?

Because we are scaling this adjusted range by the standard deviation. So, Rn divided by Sn. Now, interestingly, we will invoke this relationship. So, what happens is that if you talk about the expected value of this rescaled adjusted range, right? So, the expectation of something like Rn by Sn is proportional to C multiplied by n to the power h, okay, as n goes to infinity.

$E\left(\frac{R_n}{S_n}\right) \propto Cn^H,$ as $n \to \infty$, where C is a constant, and H denotes the Hurst exponent.

So, again, this is a large sample property, or this is more like an approximation, right? Right. Now, a couple of things to note here is that C is a constant. Right. So, capital C is a constant.

And as discussed before, this capital H denotes the Hurst exponent. So, essentially, this is the first time where we are sort of bringing this Hurst exponent in an equation form, right? So, once you define adjusted range and then rescaled adjusted range, etcetera, then this is the relationship which sort of involves that Hurst exponent. So, expectation of R n by S n—now, expectation is what? So, expectation is, on average, what happens, right?

So, in other words, the average of the rescaled adjusted range should be proportional to n to the power h, right? So, c into n to the power h. So, n is the sample size, and then this h happens to be the Hurst exponent, okay? Now, we will take a special case. So, if yt, t going from 1, 2, 3, 4 up to n, are iid, right? So, again, remember what iid means.

So, iid means identical and independently distributed, right? Okay. So, if my y t is an iid and h happens to be 0.5, right. So, in that case, h happens to be 0.5, or the Hurst exponent is exactly equal to 0.5, right. But as opposed to that, now what happens for large n?

So, again if you go back just for a second, then this is my equation in proportional terms. So, what would happen if I take a log here, right? So, what would happen if I take log on both the sides? So, essentially log of Rn by Sn for large n, of course. So, log of Rn by Sn should equal some constant alpha plus h into log n, right?

Because again, if you go back for a second, so you have this n to the power h. So, the moment you take a log, it becomes h into log n, right? And plus log c. So, log c we are denoting by alpha, by the way. So, if you go a slide up, then this alpha that you have is sort of similar to log of c. But again alpha is some constant, right, because c is a constant. So, if you take a log of that constant, it becomes alpha, but alpha itself is a constant, right. So, log of the rescaled adjusted range happens to be some constant plus h into log n. Now, the whole idea of writing this equation in terms of log is just to simplify things, because now

The Hurst exponent is simply a multiplier, sort of. So, in the earlier equation, you had n to the power h, right? So, analyzing n to the power h is sort of difficult as compared to if you have the h in terms of a multiplier, like that. And then again, interestingly, what happens here is can you can you see that this equation could be viewed as a regression equation? Now, the answer is yes.

Right. Because, let us say, if you treat this LHS to be the dependent variable and then let us say alpha to be the intercept. Right. And then H to be the slope and then log n to be the independent variable. So, I can actually visualize this entire equation in the form of a regression.

But regression between what? So, regression between the log of the rescaled adjusted range and log n. So, this is not simply the regression between, let us say, some x and some y. So, this is more specific. So, we have a regression between the log of Rn by Sn that you see here on the left-hand side and log n. So, log n is treated as the independent variable and log of the rescaled adjusted range is treated as the dependent variable, right. And alpha is the intercept; H is the slope.

$$log\left(\frac{R_n}{S_n}\right) = \alpha + \mathrm{H} * \log(n).$$

So, hence the estimate of H would be what? So, the estimate of H is nothing but the slope of this particular regression equation. So, again, just to give you an idea. So, once you collect a realization of a time series or once you collect data, I can sort of fit this regression immediately in terms of log n and log of Rn by Sn because all these values—

Rn, Sn, or log n—can be found out from that particular sample that you have, right? Because Rn was what? If you remember.

So, R_n was nothing but the adjusted range, right? So, the max of all the partial sums minus the min of all the partial sums, right? So, from that, you get R_n, and what is S_n? S_n is nothing but the standard deviation, right? And then you write this term in terms of the rescaled adjusted range and take its log. So, the entire LHS could be found out easily from the sample values that you have.

Similarly, I can easily find out log n from whatever n I have. So, n denotes the sample size. So, let us say n is 100 or n is, let us say, 50, right? So, this would be nothing but log 50 or log 100, etcetera, right? Thus, the estimate of h is nothing but the slope of the regression between the log of the rescaled adjusted range and log n.

Okay, so now, interestingly, what are the properties of the Hurst exponent, right? So, once we discuss the construction of the Hurst exponent, then we will now discuss some special cases, right? And then this sort of points to the exact problem we are trying to deal with here, right? So, if you remember, the whole problem is to capture the long memory or the persistence in the time series. So, we want to find out if the underlying time series is indeed persistent or anti-persistent, right?

So, we will discuss some special cases and then you have four of these here. So, whenever my h or the value of h happens to be between half and one right. So, firstly h is positive right here, but then h is strictly bigger than half, but then less than one then long memory structure exists ok. And by the way all these special cases can be proved using the earlier set of equations right. So, we will again not go into detailed proves here, but then you can imply some of the proves by putting some convergence ideas, right?

So, because since you have a, so H can be treated as a slope between a regression equation, right? So, H is nothing but a multiplier. So, there you can actually pinpoint to some particular values of H and then try to prove as to what happens in the future by looking at the corresponding convergence ideas. So, convergence ideas are nothing but how fast or how slow something decays as n goes to infinity or something like that, okay? So, particularly we will be more interested in what happens in a practical framework.

So, whenever H is between half and 1, long memory structure exists. Whenever H is bigger than equal to 1, so H is positive, but then it exceeds 1 also, the process has an infinite variance and is non-stationary. So, process is non-stationary and also has an infinite variance. So, on the other hand, further, whenever my h is between 0 and 0.5, right? So, whenever my h is between 0 but less than half, anti-persistent structure exists.

So, let us say for values like 0.4, 0.3, 0.2, 0.25, or 0.1. So, I can immediately say that the underlying time series is anti-persistent. As opposed to, let us say, if the value of H happens to be 0.6, 0.8, or 0.75, etc., then I can immediately conclude that there could be some long memory persistence in the underlying time series. And whenever H is exactly equal to 0.5, we saw earlier that a white noise structure exists because then it will be completely IID. It will be completely IID.

So, these are some special cases. Now, again, just to quickly recap. So, whenever H is between 0.5 and one, a long memory structure exists. So, in particular, we are more interested in this category. So, just by looking at the value of H, I can immediately see if there is long memory or not.

If my H exceeds one, then you have some problems also. Because my variance becomes infinite, and the model is non-stationary. So, ideally, this is not something that an experimenter would like to have on the first go. I mean, of course, I can adjust it later, but then if my H exceeds one, then you have some problems to deal with, such as infinite variance or non-stationarity, etc., right? Again, whenever H is less than 0.5 but still positive, right? Then anti-persistence exists.

So, again, this might be useful to tell exactly when you see some anti-persistence in the time series. And whenever h equals 0.5, then it is a complete white noise structure. Alright, now talking about characteristics of long-range persistence. So, what exactly are the characteristics of long-range persistence? So, long-range persistence is characterized by, and I think these points were discussed in the last lecture also, but this would be kind of a refresher for all of you.

One more time, right. So, instead of an exponentially decaying ACF or instead of an exponentially decaying autocorrelation function, rho k is similar to some c into r to the power k, right? And if r happens to be between 0 and 1, then the ACF decays hyperbolically with k, that is, rho k happens to be c into k to the power alpha for a positive alpha, and then alpha should be equal to 2d minus 1. So, here in this first point, just to reiterate, we are talking about two ideas. So, on one hand, what would happen if my ACF is exponentially decaying?

So, in that case, I can write down something like this structure. So, rho k is similar to some c into r to the power k. So, r has to be between 0 and 1. And whenever my ACF decays hyperbolically with k, that is, the ACF structure happens to be c into k to the power alpha. So, in the first situation, you have k in the power here. In the second case, you have k here and then to the power alpha for some positive alpha, and then we are assuming alpha to be 2d minus 1.

And the second case where long-range persistence is characterized is by looking at the rescaled adjusted range, as we discussed in today's lecture. So, the first point was discussed in the last lecture, particularly where we talked about the construction of the ARFIMA process in general and stuff like that. And the second point was discussed just a short while back in today's lecture: that the rescaled adjusted range should behave as a function of n to the power h, and whenever h is bigger than one-half. So, the rescaled adjusted range behaves as a function of n to the power h whenever h is bigger than 0.5 or h is bigger than half. And I think this is the exact thing that we discussed in the last slide also.

So, if you look at the first point here, whenever the value of the Hurst exponent exceeds 0.5 but is still less than 1, a long-memory structure exists. So, I think this is the exact same thing that is given in the next slide also. But then, these are two slightly different ways or different ideas as to how one can characterize the long-range persistence in the underlying time series. So, now we will again sort of discuss—now again, I think all these slides are sort of more like a revision. So, again, we will talk about the ARFIMA process just to close things off here.

That again, if you go back to the fractionally integrated ARMA or ARFIMA process, right, then what happens? So, in the last lecture, we saw that let y_t be generated by a stationary process. Let us say (1 - B)^d operated on y_t equals some x_t, where x_t could be written down in this structure. So, this is my ARMA structure, right? So, the phi(B) coefficient applied on x_t equals the theta(B) coefficient applied on it.

$$(1-B)^d y_t = x_t,$$

$$\phi(B) x_t = \theta(B) e_t,$$

with $-0.5 < d < 0.5$.

So, if you replace this equation in the first one, then you will get the classic ARFIMA model equation that we saw in the last lecture. And my d should lie between minus half and half. So, again, just to define what exactly the coefficient is. So, here the phi b coefficient and the theta b coefficient are the usual polynomials in terms of b. Yt is the usual white noise term.

Then we can say that Yt follows an ARFIMA PDQ process. So, I think this slide is more like a revision slide. I mean, if you are not very comfortable with what happened in the last lecture, because the idea of ARFIMA is rarely discussed. So you won't find many time series courses where people discuss ARFIMA processes. But again, in today's literature and today's requirements, the focus is more towards ARFIMA because you have some sudden shocks, as we saw earlier.

So, let's say if you are monitoring some stock price, Then how do you handle some random shocks? And the random shocks, which are not quickly reverting back to the mean. So you still have some persistence there. So capturing persistence is becoming more and more important as time progresses.

So again, let's say if you talk about global warming. So earlier, let's say 10 years back or 20 years back, The idea of global warming was there but not that predominant. But these days, you have different policies revolving around global warming. So, the idea of global warming also has some persistence in the underlying, let us say, temperature data, rainfall data, or drought data, etc.

So, understanding the practical idea of the ARFIMA process is becoming more and more important these days. So, I think this slide would be a very quick refresher about the general structure of an ARFIMA PDQ process. So, again, just to quickly summarize, this phi coefficient is nothing but the AR part. The theta B coefficient or the theta B polynomial points to the MA part. And then the value of D, let us say between minus half and half, characterizes the long-term nature of the process or the anti-persistent nature of the process.

Okay. And then again, I think this slide is also a sort of revision slide. So, let us say an Arfima PDQ process can be again summarized as this process, right? So, again, this is the exact same notation we saw earlier. Now, again, some special cases.

So, whenever D is 0, one obtains a short-memory ARMA process. Now, again, you should remember that ARMA processes are always short-memory, right? We have discussed this a number of times this week, particularly. And why exactly? Because again, if you remember, the ACF decays exponentially.

So, the correlations decay exponentially. You do not have that long-term behavior or the persistence that we want. On the other hand, what would happen if my D happens to be between 0 and 0.5? Then, in that case, the process is stationary and has long memory. And for values of D between -0.5 and 0, the process is antipersistent.

So, I think we have to focus more on this category. So, whenever D is positive but still less than 0.5. So, let us say 0.25, 0.3, or 0.4, etc. So, in that case, the process happens to be stationary and is also persistent. So, it has long memory.

When D is 0, there is not much to do here. So, you can model that using a simple ARMA process. On the other hand, if my d is negative—let us say between minus half and 0—then it indicates an anti-persistent process. So, accordingly, if you are implementing the ARFIMA process in a practical sense or on practical data, then by either looking at the Hurst exponent or the value of d, we can determine whether the process has persistence or anti-persistence, etc. Now, here we will discuss an application of the Hurst exponent.

So, we will take an example. Let us say stock price movement. Again, remember, we will redraw the same structure that we saw in the earlier class or rather the first couple of lectures this week. So, let us say if you are monitoring a stock price and you see behavior like that, there could be two cases. So, let us say if you have a random shock, then the price might move suddenly up and down erratically.

So, there will be large jumps in the process. And if you see some persistence in the time series, then it will take a while to come back to its mean. So, this indicates long-term behavior, long-term memory, or persistence. In contrast, if in the same example I redraw—let us say you are monitoring the stock price and then you have some random shock here—and suddenly you see erratic behavior or large jumps, but then the price comes back to the mean very quickly. So, after that random shock, if the price reverts back to the mean, then we will say the time series is anti-persistent.

So, in this period, we will look at how you analyze this application from a Hurst exponent point of view, right. So, suppose you are analyzing the daily closing price of a stock over the past 10 years. So, by calculating the Hurst exponent, you can gain insights into the nature of these price movements. For example, whenever my H is bigger than 0.5, let us say 0.7 or 0.8, etc., the stock price shows a persistent trend. So, this might suggest a trending market or momentum.

And then, the value of H being bigger than half we have discussed in the previous slides in today's lecture also. As for what values of H can you characterize by long-term memory or persistence, etc. So, investors could use this information to apply some trend-following strategies. So, if you have a persistent trend in the market, Then, I can apply some moving average techniques or, let's say, exponential moving average techniques or some other smoothing techniques to sort of forecast.

So, investors could use this information to apply some trend-following strategies, expecting that the positive movements will persist in the future also. Right. And if the value of H happens to be very close to 0.5, let's say 0.49 or 0.52, something like that. Now, again, the value of H need not be exactly 0.5 because you're dealing with practical data. Right.

The whole idea is that in what baskets or in what ranges of values of H can you expect the actual sample value of H to lie, right? So, H being close to 0.5 from a practical point of view could be, let us say, 0.48 or 0.52, 0.53—such values, right? So, in all such situations, we can say that the stock price follows a random walk. So, there is no observable trend, and the stock's future movements are just as likely to increase or decrease. And the third category is what would happen if my H is less than 0.5.

So, let us say 0.3 or 0.4. Then the stock price shows a mean-reverting behavior. So, after an increase, the price is more likely to decrease and vice versa, right? So, for an anti-persistent nature—after the random shock, let us say the stock price increases suddenly—then after a very short while, it will again revert back to its mean. So, this suggests that the stock is overbought or oversold and will revert to a long-term average, right?

So, this is one application as to where exactly the experimenter can apply the idea of gauging from the value of the Hurst exponent. And be it finance or hydrology, the same idea applies. So, even in hydrology, if you want to forecast down the line as to whether there is persistence in, let us say, whatever hydrological data you have, then the best way to find out is to determine the value of the Hurst exponent. And based on its value, you can then gauge whether there is some persistence or anti-persistence in the underlying time series.

Thank you.