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 29: Estimation under ARFIMA

Hello all, welcome to this course on time series modeling and forecasting using R. Now, again, if you remember, this week we are focusing on a particular kind of time series process called the ARFIMA process, right? And in the last lecture, we talked about a particular exponent called the Hurst exponent. So, by using the value of the Hurst exponent—let us say if the Hurst exponent H is bigger than 0.5 or less than 0.5—one can actually gauge the long-term memory of the underlying time series process. So, let us say, one can find out if the time series is persistent or anti-persistent, etc.

Now, again, as usual, we will start with a very short review. So, let us say these are the underlying ideas discussed in this week's lectures before today's. So, let us say ARFIMA PDQ, and even here, before defining ARFIMA PDQ, if you remember, we talked about fractionally integrated noise. So, ARFIMA 0D0, right? So, this was one idea, and here the idea is that you can find out if there is persistence in the underlying time series or not by looking at the value of D here.

So, again, the assumption is that the value of D has to be between -0.5 and 0.5, but then even in this range, when can you say that the underlying series is persistent, and when can you say that the underlying series is anti-persistent? So, all these things have been discussed in a couple of lectures in the same week, but then a couple of earlier lectures. So, ARFIMA PDQ, and again, as discussed earlier, in the last lecture we talked about the idea of, let us say, the Hurst exponent. And if you remember the construction of the Hurst exponent. So, the Hurst exponent's notation is capital H.

And how did we construct the idea of Hurst exponent is through the regression between two ingredients. So, on one hand you had the rescale adjusted range and on the other hand you had log n right and the Hurst exponent happens to be nothing but the slope of that regression. So, if y t is any realization of a practical time series. So, let us say be it some stock price movement data or let us say data coming from hydrology etcetera and so on and so forth. Then by looking at or by using the actual numerical values of the sample data or the particular realization of the time series, I can actually construct or I can actually calculate the value of the rescaled adjusted range and I can also calculate the value of log n.

Then the only thing I have to do is to create that regression between let us say log of Rn by Sn and log of n, right. And the value of h is nothing but the slope in the underlying regression. So I think these are some of the techniques we discussed particularly up to this point. And again the whole idea is to capture the persistence in the underlying time series. Now today's lecture comes with a slightly different flavor.

So I am not sure that how many of you have looked at the core structure. So down the line in some other module we will talk about the idea of spectral density in detail. But then there is a very neat looking tie between the idea of spectral density and let us say persistence in a time series. So, I thought of bringing very briefly the idea of spectral density of a time series here. So, we will start by defining what do you mean by spectral density and then eventually we will try to tie all the things down.

So, the spectral density of any time series is nothing but a way to understand how the variation in the data is distributed across various frequencies. So, to explain this more clearly, I will draw a graph. So, let us say you want to monitor the stock price movement, right? And let us say the stock price moves like that. So, you have some trend, right?

And then it can go up and again come down. So, this is just a hypothetical graph. So, let us say it behaves like that, right? Now here, can you see that I can superimpose a frequency-based approach on top of this graph? So, what does a frequency-based approach mean?

So, I can superimpose a trigonometric kind of graph. So, this could be one estimation of my underlying series. So here, we will see what would happen if I superimpose a frequency module on top of the underlying series. Now again, what does a frequency module mean? So, let us say you have this trigonometric frequency-based approach.

So this could be treated as an estimation of the underlying series. So the series that you see here, which I plotted earlier, is the actual time series process, and on top of that, I am simply imposing or superimposing a trigonometric or sort of trigonometric function on that. So this trigonometric function could be a combination of some sine waves, some cosine waves, etc. But as probably many of you might know, a combination of sine waves and cosine waves is nothing but how you capture the frequencies. Because any underlying time series process or any underlying practical data has some repetition.

So, let us say if you have seasonality. So seasonality is nothing but, let us say, you have some frequencies at some intervals. So how do you sort of gather that idea and then try to capture? Sort of turn the tables. So, I will say, to give a frequency-based approach rather than a time-based approach, right? And hence, we get something which is called the spectral density of a time series, right? So, again, just to repeat, the spectral density is nothing but a way to understand how the variation in the data is distributed across the various frequencies, right?

And this sort of helps to understand how much of the variation in a time series can be attributed to different cycles or different patterns. So, here, if you see, this would be the variation in the underlying frequency-based plot which I have superimposed. So, the idea is to capture the amount of variation in the underlying different cycles or different patterns and so on and so forth. Now, the different components of any time series—I will quickly go over what components any time series has. So, the time series is nothing but the actual data which is observed—the first thing.

The frequency stands for how often something occurs over a specific period, right? So, again, just to reiterate, if I redraw the same graph, right. So, the frequency-based approach is nothing but if you want to talk about the repetitions or if you want to capture the repetition. So, this is more like a frequency-based approach, right? And lastly, spectral density.

So, a way to analyze the time series data in terms of how much of it happens at different frequencies. So, let us say if this is the frequency-based approach, then at each peak or each trough, How much of the value are you able to capture? So, all these ideas can be sort of involved in defining the idea of spectral density of a time series. It also tells us which cycles are present and how strong or weak they are.

For example, I can redraw the same graph, but this time, rather than placing all the peaks and all the troughs equally, what I can do is, let us say, the first peak would be higher than the second peak, like that, right. It could be something like that, right. Now, again, I can superimpose a trigonometric sort of function on that because this is if you want to analyze it from a frequency-based approach. But now, for each cycle or each repetition, since the heights are different, using spectral density, I can actually estimate the heights of the peaks and the heights of the troughs also. So, in a time series, certain cycles, let us say, like seasonal changes or economic trends, are stronger or more pronounced than others.

So if you have seasonality in the underlying time series, let's say temperature data or or rainfall data, etc., then such repetitions or such cycles are much more predominant or much stronger than the other repetitions. Okay. So, firstly, why does it matter? So, why does it matter?

Because understanding patterns, right? Right. Who would not like to sort of understand the underlying patterns in the time series? So, either let's say fluctuations or seasonality, whatever. Right.

By looking at the spectral density, you can identify repeating patterns. So, let's say seasonal effects or trends, etc., in the data. Right. And the second part as to why it matters is making predictions. So, knowing which cycles are present can help in forecasting some future values.

So if you know the heights of each repetitions or if you know that what kind of cycles are present in the underlying time series, wouldn't it be easy to sort of even forecast down the line? The answer is yes. So by applying the idea of a spectral density. I can actually forecast in the future much more with a peace of mind or with much more validity by sort of gauging on the fact that how strong or how weak each and every cycle is or each and every repetition is. So, for instance, if you know that there is a strong yearly pattern in let us say sales data.

So, you have some sales data and then for an example that if you know that there is a strong yearly pattern in the underlying sales data, you can use this information to make better predictions for the upcoming year. So, the idea of spectral density is to sort of model the underlying frequencies or underlying let us say repetitions of a underlying time series. Now, we will throw in some notation and then we will try to build upon as to what exactly the spectral density function is takes the form of, right? So, here if you see, it is sort of involved, but then this capital S and then within brackets F. So, this denotes the spectral density function of any time series, right?

$$S(f) = \lim_{T\to\infty} \frac{1}{T} \left| \sum_{t=0}^{T-1} y_t e^{-i2\pi f t} \right|^2$$

- $S(f)$ is the spectral density
- f denotes the frequency
- i denotes the imaginary unit
- T denotes the total number of observations in the series

So, firstly what exactly is this? So, this is limit capital T going to infinity 1 by T and then absolute value you have a particular sum yt. Now, yt is what? yt is the underlying time series and then you have this e to the power minus i into 2 pi ft. Now, here I am pretty sure that many of you might be thrown away by looking at the intricacies of this formula.

But then I have sort of tried to explain each and every part of the formula here. So, firstly, what is Sf? So, Sf is the spectral density function. So, S is the spectral density function, and this function is a function of what? So, a function of the frequency.

So, f denotes the frequencies. So, hence, my spectral density function is nothing but a function of the underlying frequencies of the data. So, hence the notation Sf. So, Sf denotes the spectral density, and f denotes the frequency. Now, this I—so I, you see I here.

So, I is nothing but the imaginary number. So, the square root of minus 1. So, the imaginary unit—so I, then 2π, f is the frequency, and t denotes the time. And capital T denotes the total number of observations in the series. So, let us say if you have 100 observations in the series, then this capital T takes the value which is 100.

So, wherever you see capital T, for example, here—so T going to infinity, 1 by T, summation small t from 0 to capital T minus 1. So, capital T denotes the threshold. So, the number of observations you have in the time series; small t is a time indicator or time notation, and then y t. So, y t is the actual underlying data, and then e to the power minus i 2 pi f t, and then the absolute terms and then square. Right.

So, again, just to repeat. So, I am very comfortable with the fact that you have— So, this function is sort of not very easy to digest. Right. But then, the idea of how you construct a spectral density function by sort of interlacing all these other notations—

So, let us say the imaginary number 2 pi, then frequency, and then time, etc., is important. Okay. Now, the frequency f that you saw in the earlier formula is the one at which we evaluate the spectral density, and hence the spectral density function is a function of the frequency itself. And on the other hand, the exponent—so the term e to the power minus i 2 pi f t—represents the oscillatory behavior associated with the frequency f. And probably, if I am not sure how many of you are sort of mathematicians or take interest in mathematics, but this exponent has to do with trigonometric functions also.

So, I can actually write down sine theta in terms of an exponent of something, or I can write down cosine theta as e to the power of something. Or rather, I can write down a combination of sine theta and cosine theta as e to the power of something else. So, this is not a new function altogether. So, e to the power minus i 2 pi ft represents the oscillatory behavior because even sine curves and cosine curves are frequency-based. So, you have a lot of oscillations there.

So, sine curves have some oscillations, and cosine curves have some oscillations. I mean, of course, you have certain differences. But then, both trigonometric functions, be it sine or cosine, represent or have some underlying frequency aspects also. So, hence, bringing this term, which is this exponent here, is important. So, this term, which is in the exponent, represents the oscillatory behavior associated with the particular frequency f.

The exponential representation takes care of both the sine and cosine functions. So, as discussed earlier, this is nothing but a combination of sine curves and cosine curves. But then, there is one problem with the earlier estimator. So, the earlier estimator has a limit. So, again, if you go back for a second, you see that you have a limit.

So, this is more like an approximation. So, can we sort of propose an estimator to my Sf? The answer is yes. And this is called as the estimator of the spectral density or in other words it is called as a periodogram. So, this If that you see is nothing but it gives you the values or it denotes the periodogram for the underlying series.

$$I(f) = \frac{1}{T}\left|\sum_{t=0}^{T-1} y_t e^{-i2\pi ft}\right|^2$$

So, the practical estimator of the spectral density is given by the periodogram which takes the following form. So, the only difference here is you do not have a limit. So, this is not a limiting structure. So, you have 1 by capital T, then the same summation T going from 0 to T minus 1, then yt and then e to the power minus i 2 pi ft in absolute terms and then square out. So, this is nothing but the estimator of Sf.

So, if you want to write down this if is nothing but it sort of estimates my Sf or the spectral density function. Alright, now the whole build up. So, you might be wondering as to why did we transition from Arfima processes to let us say something like spectral densities and then why are we even talking about that is because the focus of today's lecture is more in terms of estimation of the parameters involved in the Arfima process. So, in the Arfima process, if you remember, you had PDQ, right? So, how do you estimate that?

So, one general notation as to how you estimate, let us say, any parameters in an ARMA process or ARIMA process is through MLE, right. I mean, of course, you have other techniques, let us say, method of moments or least squares, etcetera, but then if you remember, we had a session on estimation also. So, how do you estimate the underlying parameters of, let us say, an AR model or MA model or ARMA model, etcetera. Similarly, here, for estimating the underlying parameters of an ARFIMA process. So, I can actually use MLE, or there is a tie-up between estimation and spectral density, and hence we sort of revise the idea of spectral density initially.

So, now we will focus more on how you estimate the underlying parameters of an ARFIMA (p, d, q) process. So, estimation under the ARFIMA model. So, the first kind of estimation technique is given by Geweke and Porter-Hudak. So, Geweke and Porter-Hudak estimation. So, in short, this GPH estimator.

So, these people proposed the estimator, and hence the estimator has their name. The idea is that this uses the log periodogram of a time series. And hence, we sort of revise what you mean by spectral density, then what you mean by periodogram, etc. Because all these are the ingredients required to develop this estimator. So, this GPH estimator uses the log of the periodogram, the periodogram function of the time series.

Now, the main steps are as follows. So, let us say lambda equals 2 pi f, alright. So, lambda equals 2 pi f, right. Then compute the log periodogram of the series. So, log periodogram means what?

So, log periodogram means log of i of lambda k, right. For any given k or any order k, I can write down, I can gather the periodogram and then take the log of that. So, log of i lambda k, ok. And then the second step is to regress the log periodogram against a set of frequencies, which are lambda k, right? And then this is the regression.

$$\log(I(\lambda_k)) = a + b\log(\lambda_k) + \varepsilon_k$$

So, on the left-hand side, you have the log periodogram. So, log of i and then lambda k. And on the right-hand side, you have a plus b into log of lambda k plus some error, which is epsilon k. So, this is nothing but a regression equation, is it? So, on the left-hand side, you have the dependent variable, which in this case is nothing but the log of the periodogram. So, log of i and then lambda k, and then this entire right-hand side is nothing but you have some intercept, then slope is b, and then log of lambda k plus some random error, which is epsilon k.

Now, what? So, let us say again, where a and b are the parameters to be estimated, of course. So, again, if you go back and see the regression equation, then you have to estimate a and b. So, a is the intercept, and b is the slope, right? So, where a and b are the parameters to be estimated, and then epsilon k is the error term. The slope coefficient B is related to the fractional differencing parameter D. So, interestingly, you have a relationship between the slope coefficient B that we saw earlier and the fractional differencing parameter D. And hence, this is a tie-up between, let us say, spectral density angle and the RFEMA parameters.

So, you have a very neat-looking relationship. So, that B that we saw earlier, which is nothing but the slope of the regression, happens to be exactly equal to minus 2D. Thus, one can actually estimate the regression slope. So, I can find out something like B hat, right? So, you run the regression, and based on the underlying data, I can gather my A hat and B hat values, right?

So, I can find out the estimated slope coefficient and the estimated intercept coefficient. So, A hat and B hat. And thus, by estimating the regression slope, one can get the idea of or the estimate of d, right? Because d hat would be nothing but minus b divided by 2. So, whatever b hat you have here, you simply do minus b hat divided by 2, and this would be an estimator of my fractional differencing order, which is d.

And, of course, after estimating D, one can estimate the AR and MA parameters as usual. So, for estimating AR and MA, I can make use of MLE or Yule-Walker technique or method of moments, etc. So, once you estimate D, the only parameters remaining to estimate are the AR parameter P and the MA parameter Q. So, initially, once you estimate D by constructing that regression in terms of spectral density and then finding a slope, estimating the slope by B-hat, and then using this relationship, I can get hold of my D-hat. Then, the remaining parameters P and Q can be accordingly estimated by using, let us say, either the MLE technique or the method of moments technique, etc.

So, this is an exercise as to how one can implement the GPH estimation technique. So, now coming to some advantages and limitations of the GPH estimation. So, what exactly are the advantages? So, a couple of advantages. So, firstly, robustness.

So, the GPH method is robust in case of noise in the data. So, if you have very noisy data. So, noisy data means you have lots and lots of fluctuations in the data. So, in such a situation, the GPH method is robust. So, even if you try to estimate the underlying ARFIMA parameters using the GPH technique when the dataset is noisy, in that case, you will still get some stable estimator.

So, the values of the estimators will not diverge. So, this tendency is called robustness. And if the underlying estimators are robust, then of course it is a good sign of robustness. And the next thing is applicability. So, this technique can be applied to a wide range of time series data that may exhibit long memory behavior.

In short, I can actually apply the GPH technique to cover a lot of ground. So, if you want to estimate, let us say, data coming from different kinds of persistence, right? I mean, if let us say my Hurst exponent value is, let us say, 0.99 or 0.8 or even very close to 0.5, something like 0.6 or 0.55, right? So, by estimating the underlying parameters using the GPH technique, it proves to be robust and covers all possible scenarios. But what exactly are the limitations?

So, sensitivity to bandwidth is the first limitation. What do you mean by that? So, the improper choice of bandwidth in constructing that periodogram estimation may lead to biased estimates. So, supposedly the estimates in this case, in the GPH technique, are not unbiased. Right.

And why exactly? Because an improper choice of bandwidth in periodogram estimation may lead to biased estimates. OK. And secondly, the assumption of stationarity. So, this GPS technique—the method assumes that the underlying process is stationary, which obviously may not be the case.

Right. Because, of course, stationarity after differencing may not always hold. There might be situations where one can't even achieve basic stationarity even after, let's say, differencing twice or thrice, etc. So, under such situations, the GPH estimator may not work very well because it assumes stationarity twice. Right at the start or by applying some differencing.

So, let us say differencing once, twice, or thrice, etc. But even after you difference such series—let's say once, twice, or thrice—you still do not see any signs of stationarity; then, obviously, the dataset is highly non-stationary. And in such situations, I cannot apply the GPH estimator very accurately. So, these are some limitations of the GPH estimator. So, let us say sensitivity to bandwidth, or the second one is the assumption of stationarity, etc.

But then, at the same time, the advantages of using a GPH estimator are robustness and applicability. So, I can actually apply the GPH estimator on a wider kind of ground. So, I can cover lots and lots of different iterations or different possibilities of persistence in the underlying time series. Now, what exactly are some other estimation techniques? So, I can actually estimate the parameters using the EM technique also or the maximum likelihood estimation technique also.

Exactly how? So, such a technique involves finding the parameter values that maximize the likelihood of the observed data given the model. So, as a matter of fact, I can apply the MLE technique for estimating parameters for any underlying time series model, right? But in this case, we are focusing on an ARFIMA structure, then probably estimating using a GPH technique proves to be slightly beneficial, okay. But nevertheless, I can apply the same MLE technique for estimating parameters of the underlying ARFIMA models also.

The only thing which one has to do for obtaining the MLEs is to maximize the underlying likelihood by looking at the gathered data. So, based on the practical data that you have. The first idea is to maximize the likelihood. And of course, by maximizing the likelihood, I can then sort of get the underlying estimators of the parameters, right? So, this is one other estimation technique that I can have.

So, for our FEMA modules, the likelihood function is often derived from the joint distribution of the observed time series, right? And the second technique would be, let us say, Yule-Walker. So, Yule-Walker equations can be used, right? So, the Yule-Walker method estimates the parameters of the ARIMA models based on the autocorrelation function or ACF in short of the time series, right. So, Yule-Walker is sort of similar to the method of moments in a way.

So, what exactly is the method of moments? So, in the method of moments, you try to set up equations where you are sort of equating the population moments with sample moments, right, and then try to solve them, right. So, the Yule-Walker technique or Yule-Walker equations is nothing but the method of moments technique. Right. So, for ARFIMA models, the fractional differencing parameter can be estimated using the sample ACF, and the autoregressive and the moving average parameters can be obtained from the Yule-Walker equations.

Right. So, I can actually use the Yule-Walker estimation technique to estimate, let us say, p-hat. Right. My d-hat could be estimated by equating the sample ACF and the population ACF. So, something like my rho k and then equating this to something like rho k-hat.

Whereas I can estimate the MA parameter using some other technique, let us say MLE and stuff like that, okay. So, wherever possible, I can sort of simplify the estimation idea by applying either Yule-Walker or, let us say, MLE, etcetera, okay. And lastly, we will discuss some other estimation techniques which are alternative in nature. So, the first one is Whittle estimation, right? So, what exactly is this?

So, this method is an alternative to GPH, right, based on the likelihood of the periodogram of the time series again. But then, in this case, it uses a quasi-likelihood approach to estimate the parameters. So, quasi-likelihood is more like a non-parametric approach. So, in certain situations, if you are not able to get the likelihood function of the underlying data in a concrete formula kind of structure, then I can actually use a quasi-likelihood function. So, the Whittle estimation is sort of similar to GPH, where instead of looking at the likelihood function, it sort of focuses on the quasi-likelihood approach.

And the last one is local Whittle estimation. So, this method is a modification to the Whittle estimator and is designed specifically for estimating the fractional differencing parameter. Now, in all these estimation techniques, you might have found out that estimating my D should be the first priority, right? Because once you get D-hat, then the other orders, similarly P-hat and Q-hat, could be estimated using any technique. Let us say MLE or Yule-Walker, etc., right?

But the focus lies on how you estimate the fractional difference in order, which is D, right? And hence, I can either use GPH, Whittle estimation, MLE, or local Whittle estimation, etc. So, lastly, what is this local Whittle estimation? So, this method is a modification to the Whittle estimator and is designed specifically for estimating the fractional differencing parameter, which is D. And this focuses on a local frequency range, providing a more accurate estimate of the differencing parameter when the sample size is limited.

So, here you have a very small typo. So, this should be only D. So, let us say the differencing parameter D when the sample size is limited. So, if you do not have a very big dataset, if your sample size is limited, what it does is it focuses on some local ranges or some local frequencies. So, rather than capturing the overall frequency and then trying to put forward some trigonometric function to capture the overall frequency, it splits the data into some local frequency ranges and then tries to apply some trigonometric functions or frequency ideas on top of that. So, all these are some estimation techniques underlying the RFIMA process.

So, GPH, Whittle estimation, local Whittle estimation, MLE, Yule-Walker, etc. So, in the next lecture, it will be a wrap-up of this week's lecture, and then we will bring in some practical ideas. So, the next lecture will be in R, again the usual thing. So, we will take up a practical dataset and then try to estimate the So, firstly, we will try to model that using some ARFIMA structure.

So, we will try to estimate the underlying parameters of the ARFIMA structure. And then, eventually, we will try to forecast.

Thank you.