Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 09

Lecture 44: Numerical Examples and Further

Hello all, welcome to this course on time series modeling and forecasting using R. Now again, this week has been slightly tough because the entire idea of transitioning from a time-based approach to a frequency-based approach is not that easy, right? So again, a strong suggestion—a couple of points before we begin with anything new—is that if you want more time to revisit all the videos of this week, you should do it, right? Because the entire idea about, let's say, spectral density, Fourier transform, inverse Fourier transform, and periodogram is entirely new, right? So you might have to revisit some of the technicalities there. I mean, for example, what do you mean by frequency? What do you mean by, let's say, the relationship between e to the power minus i omega t and then, let's say, sine omega t or cosine omega t? So all these things have to be revised before you enter the last session for this week, which is the practical session, okay? But just before the practical session, we have one more session today.

And here, we will try to wrap up whatever we have been studying this week, and then the idea about, let us say, spectral density, estimating the spectral density, and towards the end of the course, I will show you some examples. So, do not worry. Now, again, where we stopped in the last lecture was we were trying to put forward some non-parametric estimations of S of omega, and one very useful estimator of S of omega is I of omega, which is nothing but called a periodogram. And what exactly is a periodogram? So, a periodogram sort of identifies whichever significant frequencies are there in the underlying signal, and we have some applications of the underlying periodogram as well.

So now again, in this lecture—in today's session—before we begin with anything new, we will revisit some non-parametric estimation framework. So, the first idea we will discuss today is called smoothing the periodogram. So, what do you mean by smoothing the periodogram? So, again, in the last lecture, we saw that the periodogram is not consistent, unfortunately, right? We saw that in the last lecture. So, again, what do you mean by that very quickly is that the variance does not converge to 0 even if you

Bring your capital T to go to infinity. So, even if you collect more and more observations, the underlying variance does not vanish, which is a bad thing, right? Because there could be some noise that is persistent, and all sorts of disadvantages can occur if the underlying variance does not sort of vanish. So, one solution in that regard is smoothing the periodogram. So, how would you smooth the periodogram to ensure that you are kind of reducing the underlying variance?

So, this is exactly the idea to reduce variance and obtain a consistent estimator. Right? The periodogram is often smoothed by averaging across frequencies. So, what I can do is look at some of the frequencies, average across all those frequencies, and put forward a single value. So, by averaging across frequencies, I am sort of smoothing out the effect of the overall variance. So, averaging always smooths out anything. So, if you are averaging a bunch of numbers, you are putting forward a smoothed value comprising all those values.

So, even within the smoothing the periodogram section, we have different kinds of smoothing. So, the first kind of smoothing is called kernel smoothing. Now again, this is slightly technical. If you want to ignore the actual technical equations here and the underlying technicalities, you can do it. But the idea should be clear as to what you mean by kernel smoothing. So, kernel smoothing means nothing but applying a weighted moving average.

So hopefully you have studied what you mean by weighted moving average long back in some previous week. So we will revisit it again here. So the idea is to apply a weighted moving average using a kernel function w. So w is a kernel function. And then we will apply that to the underlying periodogram through this equation. So, S omega hat.

So, S omega hat means we are trying to estimate that equals summation small k running from minus capital K to capital K, and then you have the kernel function here which is W of K, and then I of omega plus K delta omega. So, K delta omega. So, delta omega is a very small frequency or a small interval. So, using this equation, I can sort of smooth out the underlying periodogram by applying this kernel function. So, the extra thing here is that you have a certain summation, you have this periodogram here.

$$\hat{S}(\omega) = \sum_{k=-K}^{K} W(k) I(\omega + k\Delta\omega)$$

So, capital I stands for periodogram, by the way, is it not? So, you have the underlying periodogram here, and for each particular value of k, I am attaching this extra kernel function here, and then I am trying to take an average of that. So, this is more like a weighted moving average. Can you see that? So, this is more like a weighted moving average where weights are nothing but given by the kernel function w, ok.

So, in this period, what exactly are some of the common kernels that people use? So, these are all the common kernels. So, the first one is the Bartlett kernel. So, the Bartlett kernel means one has to apply some triangular weights, right? So, again, the idea is that if you are changing the kernel, you are essentially changing the underlying weights or rather the underlying W function, okay.

So, the first common kernel is called the Bartlett kernel. The second common kernel is called the Daniel kernel. So, the Daniel kernel means you are applying a simple moving average. So, rather than applying a weighted moving average, I am applying a simple moving average. And the third kind of very famous kernel is called the Parzen kernel.

So, the Parzen kernel means smoother weights for gradual tapering. So, essentially what you are doing is you are kind of smoothing out the weights as you progress down the line. So, it should give you a more gradual sort of tapering towards the end. Make sense? So again, I will go back to the first slide here.

So, in order to smooth the periodogram, the first option is kernel smoothing. And then we discussed a few kernels here: the Bartlett kernel, the Daniel kernel, the Parzen kernel, etc. The second approach is directly averaging over bands. So, what do you mean by that? Averaging over bands means dividing the frequency range into bands and averaging the periodogram within each band.

Again, divide the frequency range into different bands and average the periodogram within each band. So, I will give you an example. Let us say I sort of graph out the periodogram. This is exactly how the periodogram might look. And whenever you see peaks here, those are the significant frequencies, for example.

So, again, on the x-axis, you have omega, which is the frequency, right? And then, wherever you see peaks, these are the significant frequencies, for example. Now, averaging over bands means dividing the frequency range into bands. So, let's say I divide these into different ranges. So, this is my first basket, this is my second basket, this is my third basket, and then this is my last basket.

So, I am dividing the overall periodogram structure into four bands. And here, then what I will do is I will simply take the average of the frequencies in each band. So, I will average out these frequencies—these, these, and then these. So, in short, I am ensuring that I am kind of smoothing within each band. So, again, the second approach is not very difficult to understand—that you create some baskets of frequencies and then you simply take the average within each basket or within each band.

Now, the third important method is Welch's method or Welch's method. So, Welch's method divides the time series into overlapping segments, computes the periodogram for each segment, and averages them. So, again, a similar idea to the second point we mentioned just now, where we covered averaging over bands. The only difference here is that, let us say again, you have this particular periodogram—right, something like that—and Welch's method divides the time series into overlapping segments. P1 computes the periodogram for each segment and averages them.

$$\hat{S}(\omega) = \frac{1}{N}\sum_{k=1}^{N} I_k(\omega)$$

So, let us say this is P1. So, I will mention P1 here. So, what exactly is P1? P1 is the periodogram for the first segment, and similarly, this could be P2. So, P2 is nothing but the periodogram for the second segment, and so on.

So let us say yt is the underlying time series. I will divide yt into some overlapping segments. So yt1, yt2, yt3, etc. And for each of those individual overlapping segments, I will create an individual p or a periodogram. So p1 is the first periodogram for yt1.

p2 is the second periodogram for yt2, etc. And essentially then what I will do is I will take the average of all the periodograms. So, I will take the average essentially of all the individual I k omega values let us say over all k. Hopefully make sense and then I think the last equation if you read the last equation things would be more clear. I will come to that part, but this is called as a Welch's method. So, the first step is split the time series into overlapping segments.

So, y t is the actual time series is split into y t 1, y t 2, y t 3 and each segment could be overlapping. Then apply a windowing function. You require some windowing function here. These are some famous windowing functions, hamming, han. Again, this is a technical idea.

Apply a windowing function to each segment to reduce the edge effects. So, since the segments are overlapping, there could be some overlapping time series values also. So, to reduce the edge effects, wherever you have edges of the segments, you apply the windowing function. And then third step is compute the periodogram for each segment as mentioned before. So, P1, P2, P3, etc.

And lastly, average over all the periodograms. So, I can write down a simple equation like this. So, S omega hat is nothing but 1 by capital N summation k going from 1 to N. So, let us say you have N segments. So, let us say if you have N segments and essentially what you are doing is you are taking a simple average of each and every periodogram here. So, I k omega.

So, rather than forming one single periodogram which is I omega, I am dividing my time series itself into different baskets which are overlapping, creating a periodogram for each basket and then averaging over all the periodograms. So, essentially the idea is that we are trying to smooth out each and every periodogram just to ensure that the variance is reduced or we want a more consistent looking estimator. That is all. Now, the last slide here before we delve into anything new is a small comparison of estimation techniques. So, here I think things would be clear, much more clear if you sort of pay some attention to this graph here or this table here.

So, again, on the left-hand side, I am writing down the methods. So, parametric, periodogram, smooth periodogram, and the last is Welch's method. Advantages and disadvantages. So, what are the advantages of a parametric method? So, smooth, interpretable, efficient for stationary models.

But the disadvantage is it requires correct model specification, as discussed earlier. Because if you are proposing a parametric approach, then you would want to put forward a known model. Let us say ARMA, ARIMA, etc. So, this could be a disadvantage. It requires a correct model specification to be put forward.

But when it comes to the periodogram, the advantage is simple, high resolution. By the way, the periodogram is non-parametric, right? But what is the disadvantage? Noisy, inconsistent—we have seen that earlier. And hence, a solution is the smooth periodogram.

A smooth periodogram reduces noise and improves consistency. But one of the disadvantages, even after smoothing the periodogram, is that it may lose some frequency resolution. Lastly, Welch's method advantages are robustness to noise, reduced variance, etc. But one con is reduced frequency resolution due to segmentation. So, I think the entire idea about whether to go for a parametric approach or a non-parametric approach should be clear by looking at or analyzing this small table here.

So, now we are almost at the end of the theory for this week. So, we will tie everything together with a few final topics here, a few final topics. Now, some practical considerations in estimation. So, the entire last section, last session this week, and even the first few minutes of today's session have been on estimation. So, what exactly are some practical considerations when it comes to estimation?

The first consideration is the choice of windowing or the choice of the kernel function. So, a good window reduces spectral leakage while maintaining frequency resolution. So, some popular choices are, let us say, rectangular. So, high resolution but suffers from leakage, or Hamming and Hanning. So, Hamming and Hanning reduce leakage with moderate resolution loss.

So, the essential part is to choose an appropriate windowing function, let us say either rectangular or Hamming or Hanning, and to choose an appropriate kernel function. So, all these things we mentioned just a short while back, right? So, the idea about smoothing a periodogram could be made much more robust by choosing an appropriate windowing function, by choosing an appropriate kernel, etc. Okay. So, the idea is that one should be able to reduce the spectral leakage as much as possible.

And at the same time, one should be able to maintain the frequency resolution at a particular level. Okay. And the second could be the choice of bandwidth. So, what do you mean by that? So, the smoothing bandwidth in kernel smoothing controls the trade-off between variance and resolution.

So, a smoothing bandwidth controls that trade-off that we talked about in the last session between variance and resolution. So, what do you mean by that? If you pick a wider bandwidth, so wider bandwidth means less variance but more smoothing, whereas a narrower bandwidth means more variance but higher resolution. So, there is always a trade-off between how you fix the appropriate bandwidth. So, the bandwidth should not be very wide, and it should not be very narrow.

So both endpoints are kind of cons. So one should manage the bandwidth somewhere in the middle. So the choice of bandwidth is one practical consideration when it comes to non-parametric estimations. And the third could be, let's say, handling non-stationarity. So for any non-stationary process, use time-frequency methods like, let's say, wavelet transforms or short-time Fourier transforms, right?

So if you're handling any non-stationary time series, which is, let's say, entirely non-stationary, and even after differencing, it's not becoming stationary, then for such situations, you actually have to rely on some time-frequency methods. So there's a combination of both time and frequency methods. And these are some underlying techniques. So, wavelets, wavelet transforms, or STFT. So, STFT is called the short-time Fourier transform. Okay, and then sampling rate.

So, one should always ensure that the sampling frequency, which is fs, satisfies the Nyquist criterion. So, this is a very famous criterion called the Nyquist criterion. So, what do you mean by the Nyquist criterion? That the sampling frequency should be greater than 2 times f_max. So, f_max is the highest frequency. So, you look at the highest frequency, multiply that number by 2, and the sampling frequency should be at least that number. Okay.

$$f_s > 2f_{max}$$

So, ensure that the sampling frequency f of s satisfies the Nyquist criteria. So, the Nyquist criteria means that if the sampling frequency, or in general any frequency, exceeds 2 times the maximum frequency, then we are good to go. So, this is called the Nyquist criteria, okay? Now, let me again, just for a second, go back to the previous slide. So, these are some of the practical considerations when it comes to estimation.

So, handling non-stationarity, then sampling rate, right? And choice of bandwidth. And the first one is the choice of windowing, the windowing function. So, let us say rectangular or Hamming hand, and then choice of the kernel, so on and so forth, okay? All right.

So now we come to some practical applications of estimation. So, some practical applications of estimation can be one of all these. So, the first application could be in climate science. So, let's say estimating the spectral density of temperature or rainfall time series to study some seasonal and annual cycles. So, by the way, all these applications, or all these practical applications, are just to tell you the importance of estimation.

So, estimating the SDF or the spectral density. The second application could be in finance. So, let's say detecting hidden cycles in stock returns or interest rate movements, etc. The third could be in speech processing. So, analyzing voice signals to extract the pitch and tone of the speech, etc.

Or there could be an application in, let's say, vibration analysis. So, estimating the spectral density of vibrations in mechanical systems to identify faults. So, all these could be some practical applications when it comes to studying and analyzing the estimation of the underlying S of omega or the spectral density function. So, now we will focus on one particular example. So, enough of theory about estimation and so on and so forth.

So, one particular example, and then things would be slightly—probably slightly more clear here. What we will do is we will consider this time series. So, y_t equals sin of $2\pi f_1 t$ plus 0.5 into sin of $2\pi f_2 t$ plus some random error. So, ε_t is a random error as before, and my f_1 and f_2 are the two frequencies. So, I am fixing my f_1 to be 0.1 hertz. So, 0.1 hertz is a low frequency.

$$Y_t = \sin\left(2\pi f_1 t\right) + 0.5\sin\left(2\pi f_2 t\right) + \epsilon_t$$

And at the same time, I am fixing my F2 to be slightly higher, so 0.3 hertz, which is a high frequency. So, my F1 is a low frequency, F2 is a high frequency, 0.1 hertz and then 0.3 hertz. Epsilon T is a random noise with, let us say, 0 mean. So, we can assume some 0 mean for that. And now, what exactly is the goal?

So, the task is to compute the periodogram to identify the dominant frequencies F1 and F2. So, we want to compute the underlying periodogram. And the periodogram would be nothing but an estimate of the spectral density function, right? So, we would want to compute the periodogram to identify the dominant frequencies F1 and F2, okay. So, how do you do it?

So, the first step—steps in the computation—so the first step is generating the time series, right. So, you have to generate some data first. So, for that matter, let the sampling rate be delta t equal to 1 second. So, the sampling rate is 1 second, and the total time series length. So, capital T happens to be 20 seconds.

$$Y_t = \sin(2\pi * 0.1t) + 0.5\sin(2\pi * 0.3t) + \epsilon_t$$

So, these are the observations. So, small t runs from 0, 1, 2, 3, 4 up to 90. So, in that case, my underlying time series becomes nothing but that. So, y_t equals sin of 2π into 0.1t because my F1 is 0.1, plus 0.5 into sin of 2π into 0.3t because my F2 is 0.3, plus ε_t where t is nothing but this bunch. Make sense?

So, you have to fix a couple of things beforehand. So, f1, f2, Δt. So, my Δt is 1 second. So, the sampling rate is 1 second. I will be sampling at each and every second. And how many observations?

So, the total time series length, capital T, is nothing but 20 seconds. Make sense? Hopefully. Now, the next step is DFT or discrete Fourier transform. So, the periodogram is computed from the DFT.

$$Y(\omega_k) = \sum_{t=0}^{T-1} Y_t e^{-i\omega_k t}, \quad \omega_k = \frac{2\pi k}{T}, \; k = 0,1,\ldots,T-1$$

So, you have seen this even before in the last session. As to how you would construct the periodogram, right? So, the first step is to construct the DFT. So, DFT again using this formula. So, y omega k equals summation small t running from 0 to t minus 1 y t e to the power minus i omega k into t.

where my omega k is nothing but this value, which we have already seen in a prior session. So, my omega k is nothing but 2 pi k divided by capital T and k running from 0, 1, 2, 3, 4 up to T minus 1. And lastly, if you remember this formula from the last session, that is how you derive the periodogram from DFT. So, the only thing I have to do is to take the square of DFT and And then scale it down by 1 by capital T. So, essentially my periodogram, so I omega k is nothing but 1 by capital T, absolute value of DFT square of that.

Now, what exactly are the frequencies of interest? So, let us say with a value of capital T to be 20, the frequencies are nothing but small k by T and then k running from 0, 1, 2, 3, 4 up to capital T by 2. And then these correspond to all these frequencies. So, 0 Hz, 0.05 Hz, 0.1 Hz, 0.15 Hz all the way up to 0.5 Hz. Now, again remember one thing.

So, why are we only restricting up to capital T by 2? This is because we want to maintain the Nyquist criteria. So, if you remember, the Nyquist criteria was what? So, the Nyquist criteria was that we can actually only take up those frequencies which are halfway. Up to the halfway.

So, we only want to restrict ourselves to values of k running from 0, 1, 2, 3, 4 up to capital T by 2 only. And this is the bunch of all the frequencies. And the last step is to compute the periodogram. So, compute I of fk for each value of fk. So again, if you want, I can re-summarize or sort of summarize all these steps, right?

Let me go back very quickly. So the first step is obviously generating the time series, generating the data by fixing the sampling rate to be 1 second in this case and then capital T, which is 20 seconds. So my small t automatically becomes 0, 1, 2, 3, 4 up to 19. And this is the underlying time series that we have. It has two sinusoidal components.

So, sine of 2 pi into 0.1 t and 0.5 into sine of 2 pi into 0.3 t plus the random error. Then, the next step is to create or derive the DFT. So, DFT could be derived using the regular formula. And once you have the DFT, you can find out the power at each frequency using this formula by rescaling by 1 by t, taking the square of DFT, etc. Lastly, we have to identify the frequencies of interest.

$$I(\omega_k) = \frac{1}{T}|Y(\omega_k)|^2$$

So, frequencies of interest are nothing but fk equal to small k by capital T, and I am only restricting myself to 0, 1, 2, 3, 4 up to T by 2, which translates into all these frequencies. The last step is, for each and every frequency in this set, I have to compute the underlying periodogram. So, compute I of fk for each fk in the above set. So, hopefully, the steps are clear as to how you find out the periodogram if you are given a set of frequencies. So, in the same spirit, we will try to find out some frequencies manually.

So, manual computation of a few frequencies. So, let us compute the periodogram for, let us say, F1 to be 0.1 hertz and F2 to be 0.3 hertz. So, this is exactly how the Fourier transform coefficients would look like. So, at F1 equal to 0.1, I can immediately put in the equation that we saw earlier, and then y of 0.1 happens to be this guy, and at f2, which is 0.3, my y of 0.3 happens to be this guy.

Fourier transform coefficients:

$$\text{At } f_1 = 0.1: \ Y(0.1) = \sum_{t=0}^{19} Y_t e^{-i*2\pi*0.1t}$$

$$\text{At } f_2 = 0.3: \ Y(0.3) = \sum_{t=0}^{19} Y_t e^{-i*2\pi*0.3t}$$

Power spectrum: $I(0.1) = \frac{1}{20}|Y(0.1)|^2, \quad I(0.3) = \frac{1}{20}|Y(0.3)|^2$

So, this is the calculation of DFT by the way. So, I am deriving the DFT at those frequencies which is 0.1 and 0.3 and once you have that I can use the power spectrum equation for finding I at point 1 and then I at point 3 by using the earlier equation. So you rescale it by 1 by capital T, you take the square of DFT etc. So I at point 1 happens to be this, I at point 3 happens to be that.

Now the very last section we will discuss in this entire week of course in the theory session is an idea called as cross spectrum. So what do you mean by cross spectrum? So, one can actually extend from one series to two series. So, let us say X t and Y t be two stationary time series processes each with mean 0. Then the following immediate questions could be asked.

The first question is are there periodicities related to each other. So, are there some fluctuations or periodicities related to both the series X t and Y t and if yes what is the phase relationship between them. At the end of the day, the question which one can ask is that is there any frequency of a relationship between Xt and Yt? So, for this, we require the idea of cross covariance. So, I think we studied cross covariance long back in one of the earlier weeks.

By the way, cross-covariance is nothing but this: covariance between Xt and some lag of the other series, let us say Yt. So, Yt minus t. So, for creating this idea about cross-spectrum, we require some idea about the cross-covariance. And again, the moment we have the idea about cross-covariance, we can again do the entire same thing. The only difference is, rather than focusing on one series, we are transitioning to two series, which are Xt and Yt. So, the first step is to create the Fourier transform.

$$f_{xy}(\omega) = \sum_{k=-\infty}^{\infty} e^{-ik\omega}\, Cov(x_t, y_{t-k}) = \sum_{k=-\infty}^{\infty} e^{-ik\omega}\, \gamma_k^{xy}$$

So, the Fourier transform in this case is called the cross-spectrum. So, the Fourier transform of the cross-spectrum is given by fxy(ω) and then summation e to the power minus ikω. The only difference is, here we have the cross-covariance. And here, we can use some properties. So, for example, the spectrum of a sum.

So, using the cross-spectrums, we can compute the spectrum of a sum. And one can actually simplify further. And then here, if you try to simplify and put forward some assumptions, I can actually sum them up. So, if Xt and Yt are uncorrelated, it so happens that F of Z(ω) happens to be nothing but the sum of the individual Fourier transforms. So, F of X(ω) plus F of Y(ω).

$$f_z(\omega) = \sum_{k=-\infty}^{\infty} e^{-ik\omega}\, E(z_t z_{t-k}) = \sum_{k=-\infty}^{\infty} e^{-ik\omega}\, E(x_t + y_t)(x_{t-k} + y_{t-k})$$

So, this whole idea is about cross spectrum. So, again obviously technical, but I thought of putting it here as to it might be a situation that rather than focusing from a frequency domain approach on a single series, can we sort of bring in the idea of multiple series, let us say Xt and Yt. Now, the last slide is some few practical examples of cross spectrum. So, let us say identifying the leading and lagging relationships between let us say GDP growth and stock market returns to understand which variable influences the other and at what frequency. So, if you notice here, there is sort of a commonality between GDP growth and stock market returns and we want to identify the significant frequencies underlying the relationship between these two variables.

So probably the first application could be portfolio diversification. So understanding how asset returns such as stocks and bonds co-move at short or long term frequencies. And the second could be, let's say, forex and commodities. So investigating the relationship between exchange rates and commodity prices to, let's say, model some hedging strategies. So, whenever you want to focus on multiple series, we have a thing called as cross spectrum.

So, one has to analyze both the series together from a frequency based approach. So, hopefully again a small disclaimer towards the end that almost all the ideas here are completely new from what we have been studying so far in this course. So, you might want to revisit all the videos, some of the technical terminologies here, might want to revisit some of the textbooks which are given and sort of build your understanding about the entire idea about frequency domain approach before we enter the very last session in this week which is a practical session.

Thank you.