Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 02

Lecture 06: Time Series Decomposition

Hello all. So, welcome to the second week of this course on time series forecasting with applications in R. So, before we start with anything this week, I would just like to give you a short overview of what to expect in this second week going forward. So, the first idea I will discuss is called time series decomposition.

So, how exactly can we decompose any time series into its different components? And then we'll try to introduce some very basic time series models to start with. So, let's say autoregressive, moving average, or random walk, all right? And then, of course, we will end the week with another practical session in R, all right? So, the initial idea is to describe what exactly we mean by time series decomposition.

So, again, as the name suggests, we'll try to decompose a particular time series into different components, all right? And again, the idea is really simple. So, in front of you, what you see are the four different components of any time series process Yt. So, again, just to keep the same notation in place, we will use Yt as the notation to describe the underlying time series, all right. And the underlying time series could be broken down or decomposed into these four different parts or aspects, all right.

Now, we will try to elaborate more on each component here. So, the first one, which many of you might already know, is called the trend or, more technically, the secular trend. So, the more technical term is secular trend, and notation-wise, it is called Tt, okay. Now, again, just to remember, the subscripts we are using, which are t, denote the timestamp, okay. Alright, so just to describe, the first component is called the secular trend.

Now, moving on, the second component is called the seasonal component. Now, again, as described in probably one of my earlier lectures, the seasonal component is such a component that sort of repeats after some time points, right? Or you see a certain repetition in the seasonality aspect of the time series, okay. So, the seasonal component is also called seasonality in a more technical way. So, moving on, if we describe seasonality or if we are saying that a time series process is seasonal, it kind of describes the second component here, which is the seasonal component, right? Then, moving on, the third component is called the cyclical component.

So, again, we will describe the difference between seasonality and cyclicality for that matter. But nevertheless, the third important component is called the cyclical component. And then the fourth component of any time series is called the random component or irregular component. Now, notation-wise, we can see that, the trend could be described as Tt, then seasonality could be described as St, then the cyclical component can be described as Ct, and lastly, any random component or, for that matter, irregular component can be described by It. All right.

So now we will try to describe each and every component in a bit more detail so as to understand what goes inside a particular time series process or what exactly a particular time series process is composed of. Okay. So, just to move a step ahead. So, by the way, out of all the four components, the first three, Tt, St, and Ct—trend, seasonality, and cyclicality—are called the systematic components of any series. Now, again, why systematic? You may ask that systematic because these are completely due to some permanent causes and are predictable.

So, again, if you remember the Google stock price example or the Apple stock price example, or, for that matter, let us say any temperature data, right? So, what one can do is one can actually predict the Google stock price looking generally at the trend, isn't it? So, probably, I will give you a small example. So, let's say if you have a hypothetical stock example, right? And you want to plot the stock price of that particular stock, all right?

Now, looking at the general trend, one can roughly forecast it in the future. So, let's say the forecast might look like that, isn't it? Okay, so I mean, obviously, if the data is available till here, then the forecast won't suddenly be here. So, just to try to understand this idea, since you are observing a trend component, the trend component can be predictable in nature. The same thing is true for, let us say, temperature data or, for that matter, any seasonal data.

So, how does seasonal data look like? So, let us say you have repetitions. So, maybe something like that. Now, if one wants to forecast this in the future, then obviously the forecast may look something like that. So, this is a very natural way to postulate how the forecast may look.

And hence, we say that the trend component, seasonality component, or, for that matter, cyclical component are systematic components of the series. And these are completely due to some permanent causes and are also predictable. Now, the remaining fourth component, which is the random component or the irregular component, is completely due to some noise in the data. Now, if you remember vaguely, we discussed noise in one of the earlier lectures. So, what exactly is noise?

Probably, we'll discuss again very briefly that any noise is nothing but a combination of some entirely random fluctuations. So, let's say something like that. So, there's no pattern in the underlying noise. You don't see any particular repetitions in the noise data. So, the irregular component is not predictable, for that matter, right?

Since all these data points or all these observations are completely random in nature, these are not predictable naturally, all right? So, I think this slide describes all four components. So, trend component, seasonal component, cyclical component, and the random component in a bit more detail. So, if you talk about this first one, the secular trend component, which is denoted by Tt, so one can say that a trend component is a long-term movement of a time series, right?

So hopefully this is clear because if you want to describe any trend component, so trend does not happen in a really short time span, right? So, if you see a trend, hopefully the trend should last over a longer time span. So, we say that any trend component is a slightly long-term movement of the underlying time series, all right? Now, how about the second one? So, seasonality.

So, the seasonal variations or the ST component is again described as a regular periodic variation where the period of the cycle is less than or equal to 1 year. Now, again just to repeat this sentence one more time is that the seasonal component denoted by St, by the way, is any regular periodic variation where the period of the cycle should be less than or equal to 1 year. Now, again if you take the most common example which is, let us say, temperature data. So, how does temperature data look like? So, let us say there will be ups and downs depending on when there would be summers and when there would be winters, okay.

So, let us say let me write down temperature. So, temperature data at a particular location, obviously, over some time span. So, on the x-axis you have time and on the y-axis you have the underlying temperature. Now let us say that these temperatures are monthly temperatures. So, let me add something here that we are actually describing monthly temperature data.

So, in this graph here, wherever you see peaks, for example, here or here or here, those would be summer months where the temperature is slightly higher. And wherever you see troughs, for example, here or here or here, those may be winter months. So, again here you see a clear repetition which happens over this entire time span at regular intervals. So, for example, June or July of a particular year, you might see peaks, and then let us say November, December, or January for that matter, you might see troughs. And the same patterns are repeated over, let us say, 5 years, 10 years, or 15 years down the line.

So, again going back to the definition, we can describe any seasonal variation to be a very regular periodic variation where the period of the cycle is less than or equal to 1 year. So, here the period of the cycle is, in fact, 1 year because one can actually expect that the temperature which is there in June would roughly be the same in the next June, or the temperature which is there in November of a particular year would roughly be the same temperature as the next November, and so on. So, generally speaking, the repetition cycle or the period of the repetition or the period of the cycle is not more than one year. Now, similarly, talking about cyclical variations. So, what do you exactly mean by cyclical variations?

So, any cyclical variations denoted by Ct are any gradual and relatively long-term up and down movements of a series. So, now, looking at the definitions of seasonality and cyclicality, one can actually immediately write down a difference. So, what happens in seasonality? In seasonality, the period of the cycle or period of the repetition is less than or equal to 1 year, but when it comes to cyclicality, the repetition cycle is naturally more than 1 year, and generally speaking, the majority of the times the repetition cycle is at least 2 years.

Okay, so obviously later on I'll show you some particular examples, but just to describe the cyclicality idea in a bit more detail, so let's say if you think of a simple example such as a business cycle, so the cycle of any particular business for that matter, now obviously in a business cycle what happens is whenever the business is in a boom, so you can see a peak, something like that, and then let's say whenever the business is not performing that well, you might see a trough. Now again, probably the next peak might not be as high as the one you saw earlier. And again, obviously, due to the dynamic nature of the business, the next trough also need not be at the same length as earlier. So, this is a classic example of cyclicality.

So, you see some repetitions, but these repetitions are not at regular intervals that you saw in seasonality, all right. So, hopefully, this definition of cyclicality and the difference between the cyclical component and the seasonal component is clear. So, again, just to repeat one last time, seasonality should have regular repetitions, whereas cyclicality shows repetitions but not regular ones, as you see here in this graph, right. So, depending on how the business performs, you can see a peak, and depending on when the business is not performing, you might see a trough. But again, these lengths, so let us say from the recent peak to the recent trough, and then again here, so recent peak to recent trough, again, these lengths need not be exactly equal.

And lastly, talking about random variations. So, random variations, again denoted by It, are purely random or irregular. And again, as discussed earlier, this kind of resembles the noise in the data also. So, random variations as such, which are not predictable, firstly, since it's a combination of completely random observations or random movements, right? So, there's no pattern there, there's no trend there, there's no seasonality there, or there's no cyclicality there, etc.

Okay. So, again, just to give you an overview really quickly. So, the four components are trend, seasonality, cyclicality, and random fluctuations or random variations, and a mix and match of all these four kinds of constitute a particular time series process. All right. So, the next thing we will do is we will take up a small example.

So, this example is called the Nottem data set, and why Nottem? So, what exactly does this data set give you? So, the data set contains 20 years of monthly temperature measurements at Nottingham Castle in the UK, okay. So, the time span is 20 years, and the kind of data we are talking about is monthly temperatures over the entire time span of 20 years. So, starting in 1920 up to 1940.

And all these peaks or all these troughs that you see are nothing but monthly temperatures, okay? Now again, as discussed, a very short while back, since this is a classic example of temperature data, One would actually see some regular repetitions, right? So, in summer months, the temperatures would be high, and in winter months, the temperatures would be low, okay? So, all these peaks here would be the corresponding summer months of those years, and all these troughs here would be the corresponding winter months of those years, okay?

So again, I will say that this is a very classic example of seasonality. So, you have seasonality, which is present here. So, in a way, a time series decomposition exercise also tells the experimenter which component is dominant in that time series. So, I'll ask you a very simple question. So, do you see any trend here?

The answer is no. So, you only see repetitions. So, the data is not moving either upwards or downwards substantially. So, if you talk about a trend being present, the answer is no. So, I'll write down a cross.

But is there a seasonality presence? The answer is yes. So, for that matter, which component dominates this time series? It is seasonality. So, through this exercise of time series decomposition, one can actually see which component kind of dominates the underlying time series in an actual setting.

Now I'll show you the decomposition of this same example in the next slide. So, this is exactly what the decomposition looks like. The top row is the actual data that you have. The second row gives you an idea of how the seasonality progresses. Here, one can clearly see the peaks and troughs in the data and identify clear seasonality, as there are regular repetitions over the entire time span.

The third row shows the trend or the underlying trend. Here, it is evident that the trend component is not strong, as the underlying graph is not moving in one particular direction. Again, the trend is not present here; seasonality is very dominant, and the fourth row shows the remainder. By the way, the remainder is what remains in a time series after removing all other components. If you remove the trend, seasonality, or cyclicality, whatever remains is called the remainder.

Okay. Again, the fourth row shows the behavior of the remainder, and one can see that the remainder is indeed completely random. All right. This is a very simple exercise that should be performed at an early stage to determine which component dominates the time series, whether it's trend, seasonality, cyclicality, etc. All right.

Now, we will try to understand each component in more detail. We will also provide some examples. The first component is trend. Again, just to give you the definition in brief. Any long-term movement in a time series.

So, either upward or downward is called a trend. So, just to quickly chart some examples. So, let us say a trend could be, let us say, upward, something like that, or one can actually have a downward trend. Maybe something like that. So, either an upward movement or a downward movement constitutes a trend.

Now, a couple of sub-parts. So, what exactly in my trend, if you break it down into further pieces, the first point is that the trend does not contain any calendar-related influences. So, let's say, here one can't say that the trend is upward in summer or the trend is downward in winter. So, one can actually talk about a trend in a much larger time span. So, either entirely upward or entirely downward.

And the second component is without any random movement. So, a trend is predictable, as we discussed earlier. So, just to summarize, a trend is without any calendar-related influences, and a trend is also without any random movement. Now, just to understand this in a bit easier way, what exactly do you mean by a trend? So, a trend is nothing but a change in the mean level or a change in the average level.

And can you see that through both these graphs? So, let us say initially, just think in your minds that the overall mean of the data would be somewhere here. But then, as you progress down the line, as you progress down the timeline, the mean also keeps on changing. For example, then the average shifts here. Then, the next average might be here.

And finally, the next average of that sector might be here. So, can you see that the mean is also changing here? And similarly, here. So, initially, the average of all the closed observations would be somewhere here. Then the mean changes and let us say it comes down here and further down here, etc.

So, any change in the mean level or any change in the average level is called a trend. Now, the second important component is seasonality. And here, we will discuss several examples which pertain to some seasonal aspect of a time series. So, let us say any seasonality could be described as any systematic or calendar-related effect. So, we saw an example of temperature data.

So, temperature data is also calendar-related. So, summer months see higher temperatures, and winter months see lower temperatures, et cetera. But there could be many examples if you talk about seasonality. So, let's say higher sales during the festive season. So, especially in India,

if you talk about, let us say, the Diwali season or the Dashera season, then people tend to buy more diyas or people tend to buy fireworks and so on, right. So, let us say higher sales of particular items during the festive season, largely because India is entirely dependent on the lunar calendar, right. So, whenever there are festivities, there will be higher sales of the corresponding items, okay. Or, let us say, higher water consumption during summer. So, again, the seasonality aspect is there.

So, in summers, there will be higher water consumption. In winters, there might not be that much water consumption, etc. Now, underlying this aspect of seasonality, we also have an idea called complex seasonality or complex seasonal effect. So, what do you mean by complex seasonality? So, we will take up some examples.

So, the first example is weekly sales during the last week of a month or during the first week of the month. Now, think of this example for a minute in your head and then try to understand why this seasonality would be slightly complex to analyze. So, what happens in the first week of the month due to the release of all the salaries or the income of all the people or the salaries of all the paid workers that we have in India. So, due to the release of all the salaries, people tend to spend more during the first week of the month and similarly, during the last week of the month, people don't tend to spend that much due to, let's say, some financial constraints and so on and so forth, right? So, analyzing this kind of seasonality is slightly more complex than where you see regular repetitions, for example, temperature and so on, right?

The second example is, let's say, the number of trading days in a month, so different months might have different holidays or a different number of holidays. So, let us say, for example, the month of October or August has a lot of holidays, or let us say, the month of April, right? So, these are some months where the number of trading days in a month might be lesser as compared to July or as compared to September, etc., right? So, analyzing any time series and considering the seasonality aspect of the number of trading days in a month. So, due to every month having different trading days, either 20, 21, 22, or 23, depending on the number of holidays that are in the month, this is again not an exactly regular kind of repetition.

So, hopefully, the idea of complex seasonality is clear that where you cannot have exactly regular repetitions, there might be some ups and downs which are happening there. So, the idea of seasonality is slightly more complex as compared to something like temperature data or water consumption data, etc. Now, I think the next slide contains a lot of other examples where one can actually observe seasonality. So, seasonality could be actually coming from different causes, right? So, the first cause could be due to natural conditions.

So, we talked about higher water consumption during summer. So, again, this is a natural condition. So, due to more heat, people tend to consume more water, and naturally, in winter months, people tend to not consume that much water, OK? Or let us say higher milk production during winter. So, all these are due to natural conditions. Or the second idea could be due to business and administrative procedures.

So, what do you mean by that? So, a couple of examples here. So, the first one is, let us say, higher traveling during summer, where you have vacations in place, etc. Or, let us say, higher sales during the first week of the month. So, again, all these kinds of things tie to this business idea or administrative ideas.

Okay. And lastly, obviously, social and cultural behavior. So, let's say gold purchases increase on Dhanteras or during that entire season of Diwali, right? Or, let's say, Mahurat trading on Diwali, et cetera, right? So, if you understand, all these examples kind of have certain underlying seasonality in place, but that seasonality could come from different sources.

So, either natural conditions or business-related conditions or, let's say, some social conditions or cultural conditions, etc. So, hopefully, the idea of trend and seasonality is clear now. Now, moving on to the next idea: cyclicality. So, again, as described, the time span or the repetition span should be at least 2 years. So, if you look at this sentence here, the duration of these fluctuations is at least 2 years, OK.

And what do you mean by cyclical components? So, any long-term variation in a time series, which is obviously longer than seasonality, repeats in a systematic way, okay? And one can actually describe the cyclicality in a wave-shaped curve, right? So we saw the example of a business cycle a short while back. So, even there, one can actually see a wave-shaped kind of curve with some peaks and some troughs, etc., okay?

And these wave-shaped curves contain expansions and contractions. So, if you see a peak, there will be some expansion. If you see a trough, there will be some contractions, etc., okay? Now you see a slightly different kind of example, which is an actual practical example that people have studied. So, this example is the number of lynxes.

So, what do you mean by lynx? So, lynx is nothing but a particular species of wild cat. So, the number of wild cats trapped each year in Northwest Canada. And then you can see the graph in front of you. So, here, immediately one can make out that all these peaks are obviously not of the same length.

Firstly, length meaning the height. So, you can see that some peaks are higher, some peaks are lower. And also, if you see the repetitions, for example, the length from the first peak to the first trough or the length from the second peak to the second trough or the third peak to the third trough are again slightly different. So, again, the idea is that if you see cyclicality in a time series, the repetitions need not be regular; the repetitions are irregular where the peaks are slightly higher somewhere or slightly lower somewhere, right. And also, the difference between the peak and the trough may not be exactly of the same length.

Now, talking about the last component, which is the irregular component. So, the irregular component is also called the random component or residual. So, why residual? So, residual because we described a short while back that if you try to remove all the components, let us say trend, seasonality, or cyclicality from a time series, whatever remains is the residual component. So, what remains after removing the seasonality and the trend component is called the random component or the residual component.

And any residual component results from any short-term fluctuations. So, again, just to give you the same example. So, this is the example of the random component, which looks like that, okay? So, one can actually observe random fluctuations. So, the peaks are not of the same height, and the troughs are not of the same depth, right?

And the difference between the peaks and the troughs is completely random. So, this is a completely random kind of graph, right? So, any irregular component or random component results from short-term fluctuations. So, all these fluctuations that you see here are short-term. So, there's no trend, there's no seasonality, there's no particular repetition that you see.

It's completely random. So, at the last, one can actually conclude by saying that in a particular time series, any such component can be dominant. So, it can be either trend, seasonality, or cyclicality. Or if you don't have any of these, then which component is dominant? The answer is the fourth one, which is the residual, okay?

So, I can give you some examples really quickly, and then we'll stop for this session. So, the first one is, let's say, a trend. So, if you have something like that, right? So, obviously, the Tt component is dominating. Right.

Now, the second example could be where seasonality is dominating. So, we saw the example of the Nottem temperature data. So, there, seasonality is dominating. So, here, you do not see any trend for that matter. Right.

So, the component which is dominating here is St. And then, let's say, the third one is a cyclical behavior. So, let's say, a business cycle of some sort or something like that. So, here, you can see that you don't have regular repetitions, and there is no trend. So, the Ct component is dominating.

And for a structure which looks like that, where you don't see any trend, any seasonality, or any cyclicality, the random movement or the irregular component is dominating. So, in a particular time series, any component may be dominating. So, depending on which component is dominating in that series, it becomes really easy to analyze, forecast, or even model the time series in a better way.

Thank you.