Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 01

Lecture 04: Weak vs. Strong Stationarity

Hello all, welcome to this next lecture in this course on time series forecasting with applications in R. Just to summarize very quickly as to where we stopped in the last lecture. So, I guess we talked majorly about concepts like stationarity which are very, very important in analyzing any time series data. And in today's lecture we'll kind of explore more about stationarity through some examples and then exactly specify as to how one can show that a time series model is indeed stationary or not. And then we'll kind of blend in the idea of non-stationary series because as mentioned in the last lecture also, almost all the practical examples that we see around they possess some trend or some repetitions or some seasonality in themselves, making the series to be non-stationary. And just to summarize again, the last point that we talked about in the last lecture is that different types of stationarity. So, we pressed upon strong stationarity in the last lecture. And then today we will start with another kind of stationarity, which is slightly more important than strong stationarity, which is called as weak stationarity. So, what exactly do you mean by weak stationarity?

So, you have other alternative terms for weak stationarity also which are called as covariance stationarity or stationarity in a wide sense. Alright. So, either weak stationary process or covariance stationary process or in general one can say that time series is stationary in a wider sense ok. So, now, we look at the actual definition of this. So, a time series is said to be covariance stationary, if its first and second moments or first and second order moments are unaffected by change of time origin. So, one very important point to mention here is that, as opposed to strong stationarity which focus more on the overall distribution on the entire series, we don't have any such assumption here.

So, the only assumption we require is that the first and second moments should be unaffected by time. Now, what do you actually mean by first and second moments? So, that we described in the last lecture. So, the mean function or the variance function or the autocovariance function or the autocorrelation function, ok. So, weak stationary process only focuses majorly on these two or rather these four entities, and again just to summarize there is no assumption on the distribution sense.

Hence, we are putting some weaker assumption when it comes to stationarity and hence it is called as weak stationarity, alright. Now again just to summarize as to what do you mean by moments. So, that means that we should have a constant mean and variance with covariance and correlation terms being functions of the time difference only. So, what exactly you mean by that? So, let us say you take a time series process Yt and let us say you want to find out the covariance between Yt and probably Yt_{+k} something like that alright where k is the gap between the observations.

So, Yt is the observed time series at time point t and Yt_{+k} is the observed time series of the same time series of course at time point t+k. So, notation wise this would be nothing but $\gamma_{t,\, t+k}$ as per the notation.

$$Cov\,(Yt,\, Y_{t+k}) = \gamma_{t,\, t+k}$$

If a time series process is weak stationary or covariant stationary, this should only depend upon the lag or the difference between the time points which is nothing but k. So, this is the idea behind a time series being covariance stationary. So, the idea is that the final covariance that you see or the final mean function that you see or the final variance function that you see should be entirely independent of the time which is t and it should only depend on the lag which is k.

So, again I guess this slide is kind of a summary slide as to what exactly should be the assumption or the underlying assumption.

$$E\,(Yt)=\mu, \forall t$$

$$Var\,(Yt)=\sigma^2 < \infty,\ \forall t$$

$$Cov\,(Yt,\, Yt_{+k})= \gamma_k,\ \forall t$$

$$Corr\,(Yt,\, Yt_{+k})=\rho_k,\ \forall t$$

So, firstly we should require that the mean function should be completely free from t for all t. So, here since you don't see any t in the subscript so mu is a constant. Then the second assumption is the same should be true for the variance function also. So, variance of Yt should again be a constant let us say sigma square and of course, this is a very standard assumption to make. So, variance should be indeed finite ok.

Now, when it comes to covariance or the correlation functions again as described earlier. So, let us say covariance between Yt and Yt_{+k} should be completely independent of t and it should only depend on the lag which is k. And exactly same thing is true for correlation. So, correlation between Yt and Yt_{+k} should be completely free from t and it should only depend on the time lag which is k. So, if anybody tells you that I have kind of analyzed a time series which is stationary in general sense.

So, by stationary he or she actually means weak stationary not strong stationary. So, from now on, throughout the rest of the course, if we mention that a time series process is either stationary or non-stationary, we're kind of implying weak stationarity always. Okay, so I think now is a very good time to explore this entire idea about a time series process being either stationary or non-stationary using some examples. So, we'll discuss some three to four examples and then depending on all the properties we've seen in the last slide, So, again just to summarize I will go back.

So, all of these are the properties pertaining to a stationary time series. So, the mean function should be constant, variance should be constant, and covariance and correlation should be again free from the underlying time frame which is t and they should only depend on the constant k. Now how would you be able to prove such a thing? So, let us say if you are handling a very basic process, a basic time series, so how can you prove that the underlying time series is indeed stationary or non-stationary? So, I guess we will start with some very basic examples and then we'll again see a couple of advanced examples and then try to explore a bit here.

So, initially consider a time series Yt, where Yt is exactly equal to E_t. Now again I say that E_t is a pretty standard notation which people use in time series terminology. So, what exactly you mean by E_t? So, again I do not know how many of you have undergone a regression course, but in regression, the model that we have is y equal to x beta plus epsilon, right? So, then the epsilon that you have in the end is nothing, but a random error, isn't it?

Similarly, here we will say that Ets are a bunch of random errors. Ok. More specifically, we will say that Et follows some distribution, but they have to be IID. So, Et's have to be identical and independent, both, with some fixed mean 0 and some fixed variance which is sigma square. So, again just to summarize what exactly do you mean by Et? So, Et is nothing, but a vector consisting of random errors coming from any distribution it may be normal, or exponential, gamma, etcetera.

But the key assumption is that there should be IID. So, all the Ets are independent and identical. And then we can actually fix some constant mean. So, for trivial purposes, we're fixing zero as the mean and some constant variance, sigma square is the variance. Now the important question is, is this process stationary or non-stationary?

$$Yt = e_t.$$

$$\text{and } e_t \sim i.i.d.\ (0, \sigma^2)$$

So, probably we can pause here for half a minute and then you can actually think on this question as to how would you solve. Connection here is that you have to rely on all the properties we've seen on the last slide. All right, so maybe you can start with the solution. So, the first key assumption is that the expected value or the mean should be free from T, okay? So, if you find out the expected value of Et, which is nothing but the average here, so how much is that?

So, how much is the mean of Et? So, through this assumption, one can see that the mean is nothing but zero. So, we kind of assume that. So, mean is 0 and 0 is of course, a constant. So, it does not depend on t, isn't it?

So, we are done with the first property, that the mean function of E_t should be free from t. Now, firstly just one more point to add here if you may, is that we have to start by looking at the expected value of the actual time series which is Yt. But since in this case Yt is exactly equal to E_t, we can write down expected value of Et instead, which is 0. Hope this is clear. Now, talking about the second point.

So, what should be true? So, secondly, variance of C should also be free from t. So, if you write down variance of Yt which is nothing but variance of E_t because Yt is exactly equal to E_t. Now, how much is that? So, again as per the assumption that we have taken, variance of E_t is again a constant, which is sigma square. Alright. So, I can write down sigma square here and again sigma square is clearly free from t. So, variance also does not depend on t. So, I think we are done with the first two properties.

Now, the only other property which is remaining is to find out the covariance ok. So, how would you proceed with that? So, let us say we can write down something like covariance between E_t. and let us say some E_{t+k} for any k. isn't it? So, we can write down covariance between E_t and then E_{t+k} for any k of course, k should not be equal to 0 because otherwise, it will be the variance right.

Now, how much is that? So, again as per the assumption that we have taken all the E_t's are IID. So, what do you mean by that? So, all the E_t's are independent in themselves. So, if you take any two iterations of the same Et process so let us say, Et and then Et plus 1 or E_t and then E_{t+2} right or E_t and then E_{t+5}, you will get an independent structure.

So, for any k other than 0, this covariance should also be equal to 0 because they are independent. Alright, and again this 0 is completely free from t. So, this is one idea as to how one can easily prove that the underlying time series process which is E_t in this case is stationary or not. Right? So, finally, what is the answer here? So, answer is, yes, Yt equal to E_t is indeed stationary. So, do not worry.

So, we have other examples also. So, the next example is slightly involved, slightly advanced. So, consider a time series Yt where the structure is as follows: So, Yt equals Et plus one half E_t minus 1, all right. And what exactly is E_t?

So, E_t is nothing but again the same structure. So, E_t is a random error. They are IID coming from some mean 0 and some fixed variance, which is sigma square. And again, the question is, is this process stationary or not? All right.

Now, again, probably what we can do is we can again start with the first assumption. So, first assumption is, you to check for the mean function. So, mean function of Yt, which is nothing but the mean function of the process Yt. So, in this case,

$$Yt = e_t + 0.5e_{t-1}$$

$$\text{and } e_t \sim i.i.d.\ (0, \sigma^2)$$

Yt happens to be E_t plus one-half E_t minus 1, as per the equation. All right. Now again remember what the other underlying assumption is. So, again, E_ts are what? so E_ts are random errors they are IID with fixed mean which is 0. Okay. So, since all of us know that mean is kind of linear, we can actually split this into two parts, so this is the expectation of Et plus 0.5 expectation of Et minus 1

So, expected value of E_t plus 0.5 expectation of E_t minus 1 and each individual expectation has 0, as per the assumption. So, this would be nothing but 0 plus 0 which is 0. So, this way we are done with the first property. Now, talking about the second property. So, second property what we require?

We require the variance function of Yt, to be again a constant. So, variance of Yt in this case is variance of E_t plus one half E_t minus one. Now using all the properties of variance, so how do you split this variance and so on and so forth, I can write down variance of E_t Plus, since half is a constant, it will come out with a square, right? So, it will be 0.5 square variance of E_t minus 1 plus 2 times covariance of E_t comma 0.5 E_t minus 1.

Now, again just to repeat very quickly. So, we require what? So, we require the variance to be also a constant function. So, in this case variance happens to be variance of the actual model which is E_t plus 1 half E_t minus 1 which is nothing but variance of E_t plus 1 half square variance of E_t minus 1 plus 2 times covariance of E_t and 1 half E_t minus 1. Ok. Now, let us evaluate this individually.

So, how much is variance of Et? So, again as per assumption variance of Et is nothing but sigma square. So, this would be sigma square plus 0.5 square happens to be 0.25 and again variance of Et minus 1 is also sigma square because remember one thing this is true for all t. So, Et is a sequence. Et is a random sequence.

So, for any t, let us say t minus 1 or t minus 2 or for that matter t plus 1, t plus 2. So, for any subscript here, all the variances are similar and then equal to sigma square. So, I can write down sigma square here. And now lastly how about the covariance? So, since these two subscripts are different.

So, Et and then Et minus 1 and since Et is a random sequence which is independent, the covariance term happens to be 0. So, eventually how much is this? So, this is sigma square plus 0.25 sigma square which is 1.25 into sigma square. But the key assumption here or the key conclusion is that 1.25 sigma square is again free from T. So, you do not see any subscript here. So, hence we are actually done with the second point also.

1. $E(Yt)=E(e_t+0.5e_{t-1}) = E(e_t)+0.5\ E(e_{t-1}) = 0+0=0$
2. $Var(Yt)=Var(e_t+0.5e_{t-1}) = Var(e_t)+(0.5)^2\ Var(e_{t-1}) + 2\ Cov(e_t, 0.5e_{t-1})$

$$= \sigma^2+0.25\sigma^2+0 = 1.25\sigma^2$$

So, here I will give you one short homework kind of a thing. So, maybe once this session is over you can go back and try it out. So, we require one more assumption though, if you remember, which is the third assumption in our case that, you have to actually show that covariance between Yt and any other time point which is Yt plus k should also be entirely free from t. So, if you if you see k in the equation that should be fine, but the covariance or the correlation functions should be free from the subscript which is t. So, try this out.

$$\boldsymbol{Cov\ (Yt,\ Yt_{+k}) = ?}$$

So, rather than me explaining each and every point, I think it will be slightly better if you can actually write it down on a piece of paper and then try to evaluate this quantity. okay, carefully. So, again, just to repeat what has to be shown? so you have to you have to show that the covariance between yt and yt plus k is free from t. All right. Is this true or not true? all right. Okay, so we will talk about one more example which is kind of a very easy example to look at, in the first sense. And why is that? We will see.

So, consider a time series Yt, where Yt is nothing but the sum of all these errors. So, e1plus e2 plus e3 dot dot dot up to Et. So, how many errors are we adding? So, we're actually adding t errors, right? So, we're adding t errors.

$$\boldsymbol{Yt = e_1 + e_2 \ldots e_t}$$

$$\boldsymbol{\text{and } e_t \sim i.i.d.\ (0,\ \sigma^2)}$$

So, Yt is nothing but the sum of t errors, okay? And again, the same assumption with Et. So, Et is nothing but a random sequence, which is IID, where mean is zero and variance is sigma square. So, again, the question is, is this process YT stationary? Now, I think this is a very easy task to show and then you will see why.

So, the first point is you to actually show or you to actually find out if the mean function is constant or not? All right. So, how about the mean function? So, mean of Yt is nothing, but mean of e1plus e2 plus dot dot dot up to Et. So, e1plus e2 plus dot dot dot up to Et. And then here what is happening is since each of the means are zeros, right?

So, we can actually write down that mean is also zero, all right? So, eventually the mean happens to be zero, all right? Is this okay so far? So, we are done with the first point. Now, talking in terms of the second point.

So, what has to be shown? We have to show that variance function of Yt is again a constant. So, we will see what happens here. So, if you write down variance of Yt, this is nothing but variance of this sum here. So, e1 plus e2 plus dot dot dot up to Et. Now, remember one thing.

So, how much is the variance of any E I or any Et for that matter? So, as per the assumption, the variance of any Et for that matter is nothing but sigma square again. And then here, how many Et's are you adding or how many errors are you adding? You are adding t errors. So, eventually the variance of this component should be nothing but t into sigma square.

1. $E\,(Y_t) = E\,(e_1+e_2+\ldots+e_t) = 0$
2. $Var\,(Y_t) = Var\,(e_1+e_2+\ldots+e_t) = t\sigma^2$

Hopefully this is ok because each variance is sigma square each of the errors are independent. So, there is no question of covariance because covariance term would be zeros right because Ets are IID. So, covariance terms are zeros and then each variance is constant which is sigma square, but how many? So, how many errors are we having? So, we have t errors, right?

So, the variance function eventually happens to be t into sigma square. And then surprisingly, this is not free from T, right? Because you have a t component here, okay? And hence the answer is that this process is not stationary. So, answer is no, right?

And why is that? Because variance function, we've seen is not free from T, okay? All right. So, now, I think we will see one last example which is again simple and then we will move on. So, let us say you have this last series Yt which is given by a plus b t plus some random error which is Et, ok.

$$Yt = a + b_t + e_t$$

where a and b are constants and and $e_t \sim$ ***i.i.d.*** $(0, \sigma^2)$

And the assumption is that a and b are constants and Et has the same structure. So, Et is our IID with fixed mean 0 and fixed variance sigma square, all right. And then the question is, is Yt stationary or not, ok. Now again I guess handling this example is again easy you will see why. So, the first point is you have to see or you have to make sure that the mean function of Yt is free from t or not ok.

So, what exactly is this? So, mean function of Yt is nothing but expected value of the actual model that we have. So, a plus bt plus Et ok. Now remember one thing. So, expected value of any constant is the constant itself ok.

So, since a and b are constants, we can actually write down a plus bt because the expectation of the constant would be the constant itself and then how much is the expectation of Et? which is the last term here again, 0. So, this expectation boils down to a plus bt plus a0 which is nothing but a plus bt. But now the question is, whether this expectation is free from t? The answer is no, clearly, because you have a t component here. So, again, we are actually failing at the very first property itself.

So, again, hence, the answer is this process is not stationary again. And again, why is that? Because the mean function itself is not free from t. So, hopefully this is ok. So, I just left one small homework for all of you.

Again, just to repeat if you go back in the second example you would actually show whether the covariance between Yt and any other Yt plus k for any k of course which is not equal to 0. So, if you want you can mention it here that, k should not be equal to 0. So, for any other value of k other than 0 is this covariance structure free from t or not. So, through these all four examples, we've seen that whether a process is stationary or not, one can actually use a pen and paper and then try to simplify and then try to find out the answer. Now just to finish things off, so a few points regarding strong stationarity versus weak stationarity.

So, does strong stationarity imply weak stationarity or does weak stationarity imply strong stationarity? So, we have a few pointers here. Now, you should remember that strict stationarity or strong stationarity means that the joint distribution should depend on the lag, which is h or k, right. So, the joint distribution should be free from the time points t1, t2, t3 up to tk. So, the joint distribution should only depend on the lag, right.

So, one point I have to make here is that strong stationarity is putting a stronger assumption in the sense that it is required that the distribution should be free from t. or distribution should be time invariant, it should not be depending on t, whereas weak stationarity implies that the moment should be free from t. So, we are not making any assumption regarding the distribution structure when it comes to weak stationarity. Now finite variance is not assumed in the definition of strong stationarity, we do not assume that, therefore, strong stationarity or strict stationarity does not necessarily imply weak stationarity And, why is that?

So, I think some of you might have heard about a distribution called as Cauchy. So, you have a distribution in statistics called as Cauchy distribution. And it so turns out that the mean function and the variance function of the Cauchy distribution does not exist. All right. So, even if you have a structure which is Cauchy and if the time series process is strong stationary, it won't imply weak stationary because the moments are not existing.

Isn't it? Because remember what we require for weak stationary, we require that the first moment and the second moment should be free from T. But since the moments itself are not existing for a Cauchy distribution, right, so there is no question of implying weak stationarity there, all right. The third point is if you have a non-linear function of a strong stationary variable, it would still be strictly stationary, but this is not the case for weak stationarity. For example, the square of a covariance stationary process or weak stationary process may not have finite variance.

So, think of a situation just to explain this point is that, even if Yt happens to be finite variance, but the square of the weak stationary process may not have finite variance. So, if you take any non-linear function of a weak stationary process. So, let us say Yt and then for example, Yt square or something like that, but it may turn out that the variance of Yt square may not have finite variance. So, since it does not have finite variance, the second moment is ruled out ok. And last point is weak stationarity usually does not imply strict stationarity. Also, as higher moments of the process may depend on time t.

So, these are some properties or these are some pointers regarding when can strict stationarity imply weak stationarity or vice versa. Now coming from that point, we'll discuss a very important thing in the next slide, which is this one, that if the underlying time series process Xt, so again, I've used a slightly different notation. So, Xt is nothing but the same Yt notation that we're following. So, if the underlying time series Xt is a Gaussian time series. So, what do you mean by Gaussian time series?

So, Gaussian time series means that the underlying distribution is a normal distribution. or the underlying distribution functions are all multivariate normal. So, if you treat a vector right. So, if you impose this extra normality assumption then in that case weak stationarity also implies strict stationarity. So, I guess this is one very strong way where weak stationarity implies strong stationarity and why is that because all of you know that for a multivariate normal distribution or for that matter for any normal distribution

The first two moments completely characterize the distribution, right? So, if you want to define a normal distribution to somebody, then you only have to specify what the mean is and what the variance function is, alright. So, hence if the underlying distribution is Gaussian, which is again a normal distribution. So, Gaussian means normal, by the way. Then we will say that weak stationarity also implies strict stationarity, ok.

So, hopefully this is ok. And now kind of we will summarize as to what we mean by non-stationarity. So, now we are shifting slightly from stationarity to non-stationarity because I guess non-stationarity is more important because almost all the practical examples exhibit some non-stationarity. So, there could be either trend or seasonality or some pattern here and there. So, small definition is that if a process lacks stationarity or statistical equilibrium, it is called as non-stationary.

And following are very important three points which amount to any time series being non-stationary. And what are they? So, firstly, trends. If a time series has some trend, let us say upward trend or downward trend, then we can say that the time series is not stationary. Or the second point is seasonality.

So, if you have some repetitions. due to some seasonal variations or cyclicality right. So, again maybe something like that right. Then again, we can say that the time series is not stationary. And the last aspect is changing variance.

So, changing variance or in time series terminology we call this a heteroscedastic in nature. Heteroscedasticity. So, what do you mean by that? So, I will give you very quick example here. So, let us say if the variance of the process is changing.

So, maybe something like that. So, here can you observe that the variance initially is small and as you go down the timeline the variance is increasing. Alright. So, if you immediately observe any trend in the underlying time series or seasonality or changing variance, then you can actually safely say that the underlying time series is not stationary. Okay.

Thank you.