# Time Series Modelling and Forecasting with Applications in R

**Prof. Sudeep Bapat**

**Shailesh J. Mehta School of Management**

**Indian Institute of Technology Bombay**

**Week 04**

## Lecture 19: Diagnostic Checking -2

Hello all, welcome to this course on time series modeling and forecasting using R. Now, again, just to give you a quick overview of where we are standing with respect to the course material as of now. So, I think in the last lecture, we started with this entire area about diagnostic checking, right? So, let us say you create some model, you identify the optimal orders of the model. So, let us say an ARMA model with orders P and Q, and then in the last lecture, we saw how you identify the optimal orders based on some probably information criteria. So, let us say AIC, SBC, HQIC, etc.

And then, I think the next step that we covered was model estimation. So, once you identify the correct orders, then how do you estimate all the underlying unknown parameters in the model? So, let us say if you are taking an ARMA model, then what exactly are the unknown parameters? So, let us say the set of all phi coefficients, then the set of all theta coefficients. And at the end of the day, you have that sigma square E, which is nothing but the variance of the error term.

So, this is a collection of all the unknown parameters which need to be estimated using either, let us say, the maximum likelihood approach or something like the method of moments approach or probability squares approach, right? And then, in the last lecture, we talked about model diagnostic checking. So, once you are through with, let us say, fixing the orders and model estimation, then how do you basically ensure that all the model assumptions are met or not, okay? And I think in the last lecture, we started somewhere here. So, we were discussing how you basically detect serial correlation, right?

Now, just to give you a very quick overview of what exactly you mean by serial correlation. So, let's say if you have a random graph or, let's say, a time series process which looks something like this, right? So here, we can see that you have some serial correlation because any two consecutive points, right, are either above the zero line or below the zero line. So, let us say if you have a zero line, something like that, then let us say if you take any of the chunks.

So, for example, this one or probably that one or the last one. So, in each of these individual chunks, you can see that all the observations are kind of located on the same side of the zero line. So, in a way, any two consecutive observations or any two consecutive lags are kind of correlated, and this can be seen throughout the timeline. So, again, on the x-axis, you have time. So, let us say as you go down the timeline, the same identity or the same kind of categorization can be observed.

And I think in the last lecture, we talked about one particular hypothesis testing procedure called the Box and Pierce test. And there, we talked about the corresponding test statistic and then, based on the test statistic, how you come to a conclusion and so on and so forth. Now, just to complete this idea of serial correlation, the first thing we will do in today's lecture is kind of discuss another hypothesis testing procedure which was given by Ljung and Box. So, again, these are two different people who came up with this modified test kind of jointly. So, hence the name Ljung and Box modified test.

$$Q_{LB} = n^2 \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

Now, under here, you have this slightly complicated-looking test statistic, but nevertheless, it is called the QLB. So, LB stands for Ljung-Box. And then, what exactly is this test statistic dependent on? So, you have n squared and then a particular summation of all these squared estimated correlation terms, right? So, these are squared estimated correlation terms because you have rho hat squared, right?

Depending on, obviously, how many lags you are taking and then divided by n minus k. So, just one point to make here is that if you look at this structure of the test statistic, it is slightly different from the one we saw in the Box-Pierce test, right? So, you have a slight modification here, and hence the name Ljung-Box modified test, okay. Now, what exactly do you mean by this H that you see in the formula? H denotes the number of lags being tested.

So, let us say 10, 15, or let us say 20. So, depending on how many lags or up to what lag you want to test for serial correlation, you can actually fix the value of H here in the formula. And then, as usual, N denotes the sample size. And once you get hold of any test statistic, for that matter, regardless of whatever hypothesis testing procedure you are performing, If the test statistic happens to be bigger than the tabulated chi-square value, then you have to reject the null, right?

So, this means that if you are rejecting the null, there is some autocorrelation among the residuals, okay? Because remember that the null hypothesis is that there is no serial autocorrelation or there is no serial correlation among any of the lags. And then, obviously, the alternative hypothesis is that there exists some serial correlation among the lags. So, if you are not able to reject the null, then we can say that there is no dependency or there is no serial correlation. On the other hand, if you are able to reject the null, then this means that there is some autocorrelation among the residuals among some of the lags.

And one possible solution is that the model needs to be rechecked. Now, again, what do you mean by that? So, again, from the last lecture, if you remember, the model needs to be rechecked means that probably you may add one or two lags in the AR component or the MA component of the corresponding, let us say, ARMA model or ARIMA model, etc. So, I think this is broadly speaking how you detect serial correlation or dependency among certain lags of the residuals. Now, again, before we proceed to the last item in diagnostic checking, I will just quickly remind you of the fact that we are doing all these procedures, for example, let us say serial correlation or checking for normality or checking for, let us say, independence, or the last thing we will do is checking for constant variance.

So, all these ideas are implemented on the residuals of the model and not the actual time series process. So, just to make sure that you are clear on this idea here, right? So, you fit the model using some technique, then you get hold of the residuals, which is nothing but y minus y hat kind of a situation, and then once you get hold of the corresponding residuals of that model fit, all these diagnostic checking entities are applied on the residuals of the fit, okay? Okay, so I think the last thing that we will cover is how you detect changing variance. So, changing variance is again a problem, as discussed even earlier in some of the earlier lectures.

And by the way, this is a very time series technical term for changing variance, which is called heteroscedastic in nature or heteroscedasticity. Okay. Now, again, if you have a constant variance, then the term is homoscedastic, okay. So, for example, it is asking here, right? So, heteroscedasticity implies changing variance.

So, what will be the term for constant variance? So, the term for constant variance is nothing but homoscedastic. So, we will say that a process is homoscedastic in nature or there exists some homoscedasticity if the variance is kind of constant over the entire time span, okay. Now again, before you proceed with any testing of hypothesis kind of a situation or something like that, I will again quickly remind you of certain situations where such a thing can happen. Now again, just to quickly remind all of you that detecting heteroscedasticity is for the residuals only or residuals again.

So again here, we are not concerned about the changing variance which is seen in the actual time series process. But let us say you fit the process in some model, then you get hold of the residuals, and then all these diagnostic checks have to be applied on the residual data. Now let us say, for some reason, the residuals are behaving something like this. So here, clearly, we are seeing a changing variance pattern because initially, the variance is small in this region. And then, as you go down the timeline, the variance is kind of increasing.

So here you see a fanning-out kind of pattern. Right. And exactly opposite to this can also be observed, let us say something like this: initially, you have quite a lot of variances, but then as you go down the timeline, the variance is kind of reducing, OK? Probably something like that, OK. So, even though initially you have large variance, but then as you go down the line, the variance is decreasing. So, this is more like funneling in, right?

So, fanning out and then funneling in kind of an effect, OK. Or, for that matter, there could be some other very complex kind of pattern where you see some changing variance happening in, let us say, localized neighborhoods. So, what do you mean by localized neighborhood? So, let us say initially the variance kind of oscillates in this manner, and then suddenly there is an increase in variance, and then probably again the variance is kind of again in a very tight band, something like this. So, rather than kind of variance being increased or decreased like an overall variance,

So, you see some localized behaviors of kind of fixed patterns among the changing variances. But now, nevertheless, if the variance is changing, it becomes difficult to kind of monitor or kind of model the data using a single model, and then we have seen that before. So, hence this is again a problem; it requires some diagnostic checking. So, now again, what do you exactly mean by this in terms of, let us say, if you want to use a technical term for this? So, let us say it occurs if the errors are having a non-constant variance structure or, in other words, variance of the error terms.

So, the variance of E_t happens to be sigma squared t. And then, clearly, the sigma squared T, since it depends on T there, it must be changing as T changes. Okay. Or otherwise, if you simply write down, let us say, the variance of E_T to be some constant. So, let us say sigma squared.

Then here, since sigma squared does not depend on that T, we will say that the variance is kind of constant. All right. So, hopefully, the idea of changing variance is clear. And then, the idea as to why it is a problem is clear. Okay.

Now, how do you deal with it? Or, for that matter, how do you detect it? So, the first very important idea is to create some ACF and PACF plots, right? So again, ACF is the autocorrelation function, and then PACF is the partial autocorrelation function. And I think a couple of classes back, we talked in detail about how you plot these two and then how you read from these two plots and so on.

But here, a very important kind of comment is that rather than creating ACF and PACF plots for the actual data, We have to create ACF and PACF plots of the squared residuals and not the actual residuals. So, you have to create ACF and PACF plots for the squared residuals. Now, of course, one can actually analyze ACF and PACF plots of the actual residuals also, but creating ACF and PACF plots of the squared residuals is much more informative. And we will see exactly why, okay.

So, again, just to summarize, rather than creating ACF and PACF plots of ET, we have to actually go for creating ACF and PACF plots for something like ET square. And why exactly? Because one very important assumption in our case is that the expected value of all the errors is nothing but 0, right? So, this is an assumption. So, the expectation of all the error terms is 0. And since the expectation is 0, if you talk about the variance of ET, because here, remember that we are more concerned about changing variance and not changing mean, right? Because the mean is 0.

But we are more focusing on how the variance structure changes, okay. So, if the mean is 0, the variance becomes nothing but the expectation of ET square, okay. This is a very standard formula, right? Because the variance of any random variable is nothing but the expected value of X square minus E of X whole square, right? But then, in this case, E of X is simply 0.

So, the variance of E_t boils down to the expectation of E_t squared in a way, right? So, would it not make more sense to focus more on the squared residual term rather than the actual residual term because variance is directly linked to the square of the corresponding residuals, okay? So again, just to reiterate the same idea, you have a small comment here that since all the errors have a 0 mean, the variance is defined by the expectation of the second moment. So, the second moment is nothing but E_t squared and the expectation of that, which are nothing but the squared residuals. Thus, if the variance of errors is indeed constant, then what should you observe?

So, the ACF and PACF plots of the squared residuals should be within those 95% bounds or 95% limits. So, in a way, if indeed the variance happens to be constant, then you should probably see something like this. So, let us say these are the 95% bands. So, probably apart from the first correlation, basically all the further correlations should be within the band.

So, this and then again this should be true for both ACF as well as PACF, by the way. So, again in the PACF, we have the same structure, right? So, you have these 95 percent limits, and then again if all the correlations apart from, let us say, the first one are in between the bounds, we will say that they are not significant, basically. So again, if you remember from one of the earlier lectures, if all the spikes apart from the first one are in between the bounds, we kind of conclude that all the correlations at those particular lags are not significant. And hence, there is no question of basically changing variance because you have a stationary kind of a residual series.

Because if you are observing something like this in both the ACF and the PACF plots, you right for the squared residuals, then we can actually assume that the squared residuals have a stationary structure. And whenever any model or any process has a stationary structure, the variance has to be constant. Okay? So, what do you again mean by stationary structure?

So, probably if you are observing something like this in both the plots, and then you create a simple plot of, let us say, t versus something like et square, then it should kind of resemble a completely random and a stationary structure where the mean is constant, the variance is constant, etc. So, in a way, the ACF and PACF plot of the squared residual is a very good kind of visual check to ensure that the variance is indeed constant or is it changing. But, of course, we will not stop just by looking at some visual checks. We, of course, have some formal hypothesis testing procedures also. But then, just before covering those, we will cover a small section about what could be the outcomes of changing variance or what could be the certain outcomes of heteroscedastic nature, okay. So, these are all the outcomes. So, let us say parameter estimates would be unbiased, but not efficient, okay. So, if the variance is changing, even if the parameter estimates are unbiased, they may not be that much efficient, okay. And the second one is, one has to use generalized least squares or weighted least squares for estimation.

So, rather than using something like ordinary least squares or OLS, which kind of breaks down here because variance is changing, right? So, you have to adopt some other techniques. So, let us say generalized least squares or weighted least squares, OK? So, probably we will talk about this idea in a slightly later lecture, OK? But probably just to give you a very short overview of, let us say, how would you apply WLS.

So, WLS stands for weighted least squares. Now again, the idea is if you can just visualize in your heads that if the variance structure is kind of changing, for example, like that. So, you have to control the changing variance somehow. So, one very important way here is that, let us say, you put some larger weights for these observations because here the variance is small. So, let us say you put some large weights to all of these residuals, large weights, and then wherever you see larger variance, you put small weights.

So, for example, here you put small weights. So, again, if you kind of combine this idea, then at the end of the day, you are kind of balancing the entire variance structure. So, if you are putting smaller weights to larger variance and larger weights to smaller variance, then are you not balancing out the entire variance structure, OK? So, this is an important idea or important solution when it comes to conquering or kind of controlling for that changing variance idea. And now the third one is that the estimate of the variance is also a biased estimator.

So, unlike if the process is stationary where the sample variance is unbiased for the population variance, but then here in this case, if you estimate the variance structure, that will not be an unbiased estimator, but it will be a biased estimator, which is a problem. So, thus classical testing procedures are kind of invalid because you do not have unbiasedness, you do not have efficiency, and stuff like that. So, these are some kind of important outcomes if the variance structure is indeed changing. Okay. So, I think now we will discuss two important hypothesis testing procedures when it comes to kind of controlling or kind of detecting changing variance.

Alright. Now, the first one is called White's general test. So, again it is given by a person called White, of course. And again, let us say what exactly is a null hypothesis. So, in both the testing procedures, in fact, the null hypothesis is kind of similar.

$$H_0: Var(e_t) = E(e_t^2 | Y_{t-1}, Y_{t-2}, \dots) = \sigma_e^2 \text{ (constant)}$$

So, let us say H naught is, if you talk about the variance of the residuals. This is nothing but the expectation of E t square given all the data up to, let us say, t minus 1, right, which is nothing but sigma square E, right, which is a constant because sigma square E does not depend on t, ok. So, in other words, H naught is nothing but that the variance is constant, basically, ok. So, in other words, this means that the null hypothesis is the variance is constant. So, the variance is not changing, ok.

Now, again here, if you clearly see that we do not have any expectation of the ET term because the expectation of ET is 0 as per the assumption. So, the variance kind of depends only on the expectation of the ET squared term. Okay, so let us say after estimating all the parameters, so let us say you are working with some ARMA model, AR model, ARIMA model, ME model, whatever, but then let us say you come up with all the estimated parameters, you plug in all the estimated parameters in the model, and then you can actually obtain the residual and its square, isn't it? So, let us say if the actual time series process is y t, then you apply some estimation technique, get hold of y t hat, and then these are nothing but my e t hat values, which are residual values. So, from these e t hat values, if you simply square these, you can actually also get e t hat square values.

So, after estimating the parameters, we get hold of e t hat and e t hat squared values. Now, here under the Weitz general test, the idea is kind of not that difficult. So, you construct this following artificial regression. So, here this is a slightly complicated model maybe at first glance, but we will try to explain what is going on here. So, how exactly is this artificial regression? Because you are kind of regressing some variable on a few of the other variables, and hence it is a regression, and then again, this equation on the right-hand side contains a bunch of unknown parameters.

So, again just for a second, if you kind of look at this model, this should resemble a kind of linear regression or, for that matter, some non-linear regression kind of structure, okay. Now, what exactly goes inside this regression is that now we are focusing more on E t squared values, right. So, we are kind of assuming that the E t square values are the dependent variables. And then here, what exactly are the independent variables in this regression? They are nothing but the historical values of the time series itself. So, let us say y t minus 1, y t minus 2, along with let us say their square term.

So yt minus 1 square, yt minus 2 square, right along with any of the interactions, right. So, by the way, these kinds of terms where you have a product of two lags are called interaction terms. So, y t minus 1 into y t minus 2, right. One important point here is that why do you want to include all the squared terms here? Because we are focusing more on the E t squared. And then E T square should come or should contain some notion of squared time series processes also.

$$e_t^2 = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \gamma_1 Y_{t-1}^2 + \gamma_2 Y_{t-2}^2 + \cdots + \delta_1 Y_{t-1} Y_{t-2} + \cdots + u_t$$

So, Y T square or Y T minus 1 square, etc. So, in order to construct this artificial regression, we are kind of assuming that E T square is a dependent variable which is kind of dependent on all these other values. So, y t minus 1, y t minus 2, y t minus 1 square, or for that matter, all the interaction terms y t minus 1 into y t minus 2, etc. And then, lastly, you have this error term. As you have in any other regression model.

And then you have this intercept. So alpha naught. And again here. So let us say alpha 1, alpha 2, gamma 1, gamma 2, delta 1 are all the unknown parameters. Okay.

So, what next? So, once you create this artificial regression, then obviously if the variance is kind of constant, right, or under the homoscedastic nature. The homoscedastic case implies that all these coefficients should be equal to 0, isn't it? So, again, you can go back a slide and then from this regression, you can actually sense. So, if all these coefficients are 0, by the way, alpha 1, alpha 2, gamma 1, gamma 2, delta 1, then my E t square is kind of independent of time.

$$\alpha_1 = \alpha_2 = \cdots = \gamma_1 = \gamma_2 = \cdots = \delta_1 = \delta_2 = 0$$

$$nR_{\hat{e}_t^2}^2 \sim \chi_m^2$$

Where $m$ is the number of variables in the artificial regression, except the constant term

So, my E t square term is independent of time and hence stationary, basically. So, under the homoscedastic case, all these intermediate coefficients should be equal to 0, which sort of boils down to this particular test statistic. Okay. So, you basically look at n into r square of et hat square. Okay.

Now, by the way, what exactly is r square? So, this r square is nothing but the coefficient of determination of the artificial regression that you have created. Okay. So, the coefficient of determination of the artificial regression that you have created and then this should actually follow a chi-square distribution. Okay.

So, this is the test statistic, and again, once you get hold of a test statistic, we kind of test the hypothesis in a usual manner. So, now again, one more point here is that where n is the number of variables in the artificial regression. So, how many variables are there in the artificial regression is nothing but given by M except the constant term. Now, again, the same kind of conclusion. So, you look at this test statistic.

So, either if the value of this test statistic exceeds the tabulated chi-square M value, then you reject the null, or the other way to conclude is based on the p-value basically. So, this is pretty much the Weitz general test, and then, on the other hand, you have another test which is called the Breusch-Pagan. Okay. So, again, the idea is kind of similar. So, again, under the null hypothesis, you have the same kind of structure that variance should be constant.

$$H_0: Var(e_t) = E(e_t^2 | Y_{t-1}, Y_{t-2}, \dots) = \sigma_e^2 \text{ (constant)}$$

Okay. Now, here, the idea of how you construct that intermediate regression is slightly different. So, here, we are not basing, if you look at this closely, we are not basing our E t square hat values on any of the squared terms. So, can you see that? So, in the Weitz test, we are kind of basing those terms on the y t minus 1 square or y t minus 2 square terms or y t minus 1 into y t minus 2 terms, right?

$$\hat{e}_t^2 = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_m Y_{t-m} + u_t$$

But in this regression, you do not see any such terms. So, this is more like a simplistic-looking regression term. So, alpha naught plus alpha 1 yt minus 1 plus alpha 2 yt minus 2 all the way up to alpha m yt minus m plus the error term. Ok. And again, the same kind of story that tests if all the slope coefficients are 0 or not.

$$\alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

$$nR^2_{\hat{e}_t^2} \sim \chi^2_m$$

Where $m$ is the number of variables in the artificial regression, except the constant term.

So, again, if all these alpha 1, alpha 2 up to alpha m happen to be 0, then my E t square hat term would be kind of constant, right, and then the square residual would resemble a stationary kind of series, ok. So, under both these tests, the idea is kind of similar, right. Essentially, what you are testing for is whether my intermediate slope coefficients, so alphas or gammas or deltas, are equal to 0 or not, basically. Yeah. So, this is again under the homoscedastic case that if all the coefficients happen to be 0, then you have a constant variance case, ok.

Then you have a constant variance case. And so, again, under the Breusch-Pagan test, you have a similar kind of story that this is the test statistic under Breusch-Pagan where again m is the number of variables in the artificial regression except the constant term, and again, the same conclusion can be drawn. So, if you look at the p-value and if the p-value happens to be less than alpha, the corresponding alpha, then you reject the null or otherwise you fail to reject the null, ok. Alright, so what exactly could be the consequences of heteroscedasticity, right? So, once you encounter changing variance or once you encounter heteroscedasticity in the time series process or the underlying residuals, what could be some of the consequences of that?

So, if there is heteroscedasticity, the error variance is not constant, right? And it is changing over time, right? So, I think we discussed this a lot in this lecture: the error variance is not constant and is changing over time. Thus, we need to model the volatility of the underlying process. So, in time series literature, you have a thing called how volatile the time series process is or if there is volatility in the model.

Or is there volatility in the underlying time series process? And one requires some sophisticated time series models for that, which are called, let us say, ARCH or GARCH. So, what is ARCH? So, ARCH is autoregressive conditional heteroscedasticity, and GARCH is generalized autoregressive heteroscedasticity, alright? Now, again, I am missing 'conditional' here.

So, both are conditional, by the way. So, ARCH is simply autoregressive CH, and then GARCH is generalized autoregressive CH, okay? And probably we will discuss both these models, ARCH and GARCH, in very much detail at a slightly later stage of the course, right? Where we will talk more about how you model volatility, right? Or rather than modeling the actual mean or something like that, okay? So, more on these models slightly later.

So now, in the next 2-3 minutes, we will quickly discuss a practical approach or a practical example and then combine all that we have discussed. So, again, you must have probably seen this data sometime back in one of the earlier lectures. So, what exactly is this? So, this is the closing stock price of Google for 1000 consecutive trading days between, let us say, 25th February 2013 and 13th February 2017. So, roughly 4 years of data.

Okay. And this is exactly how the time series of the Google stock price looks like. So, you have some trend, right? So, you have some sharp behaviors. For example, here you can see that the price suddenly jumped, right?

In this region, okay? So, the variance might not be constant again. So, we have to delve deep into this example. Now, again, all these kinds of plots and pictures can be easily obtained in R. So, probably, we will discuss in depth in the next lecture, which is a practical one, okay? Okay, so by the way, here at the top, you can see a simple plot of the residuals.

Okay, so residuals using a naive method. So, this is exactly how the residuals behave. So, here again you can see a sharp change or upward behavior or sharp jump in the residuals. But again, if you ignore this, apart from that sharp jump, all the residuals are kind of resembling a stationary series. And again, if you draw the ACF plot of the residuals, then all the correlations are inside the band.

So, it means that the correlations are not significant. Now again, if you talk about normality, so here you see that you have an outlier here, right? But apart from this outlier, otherwise you have a symmetric kind of curve, isn't it? Okay, so if you ignore this outlier, and probably this outlier is again clearly because of this jump that you see here, alright? So, if you ignore that outlier completely, then otherwise you have a kind of normal-looking structure.

So, essentially, we do not have any major problems when it comes to, let us say, the behavior of the residuals or the ACF plot or normality, etc. So, again, this is a normal QQ plot. So, the normal QQ plot now again shows the outlier here, but again, if you ignore that, then almost all the points kind of fall along the straight line, which kind of ensures normality in a sense. And then lastly, if you talk about the Box-Pierce test or the Box-Ljung test. So, in both these tests, you can see that the p-value is bigger than alpha.

So, 0.3886 and then 0.3551. So, since the p-value is bigger than alpha in both cases, we fail to reject the null, right? So, we fail to reject the null, which means there is no serial autocorrelation whatsoever. So, what could be the conclusion? So, the conclusion is, apart from that jump,

So, you do have some outliers. So, how do you deal with outliers? That is one situation. But let's say if you simply ignore the outlier. Because here, if you ignore the outlier, you are only ignoring just one observation, pretty much, right?

So, let us say the best thing to do here could be ignoring the outlier and then refitting the model, okay? So, if you ignore the outlier, other than the outlier, you do not have any major problem. So, both the serial correlation checks are kind of matching. And again, if you go back just a slide, then here you can see the normality, apart from this outlier, is kind of met, right? And here, let us say the behavior of residuals, apart from this sharp jump, is kind of looking stationary, right?

There is no serial correlation in the ACF plot, and from the histogram itself, it looks normal. So, it kind of checks all the boxes as we wanted. So, I think in the next lecture, we will do some practical things in R and then just sort of combine this entire idea about diagnostic checking in general. Thank you.