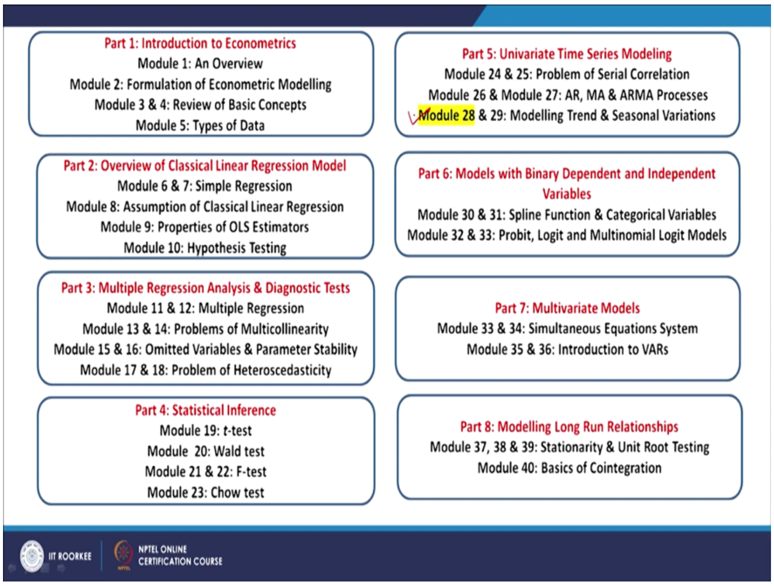
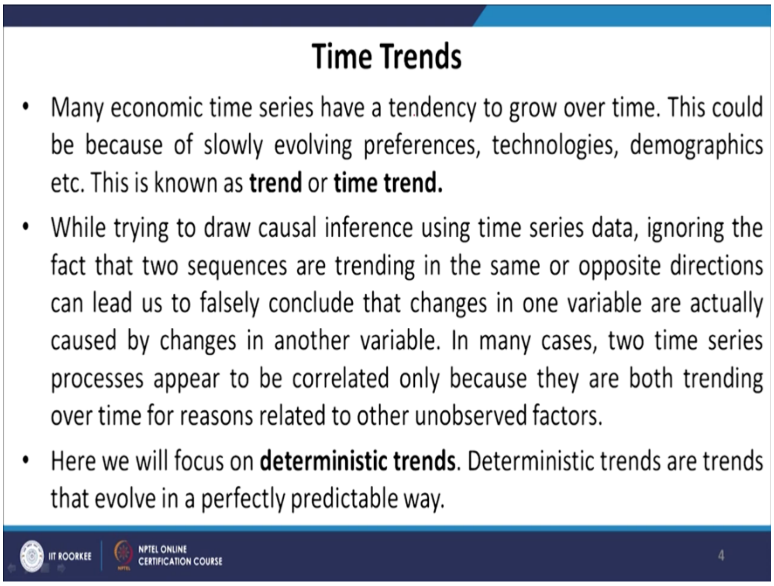
Econometric Modelling
Professor Sujata Kar
Department of Management Studied
Indian Institute of Technology Roorkee
Lecture 28
Modelling Trend & Seasonal Variations - I

(Refer Slide Time: 00:30)



Hi and welcome back to the course on econometric modelling. This is module 28 and it starts discussing the last topic under the univariate time series modelling. Here, I am going to deal with modelling trend and seasonal variations. So, two modules would be dedicated to the discussion on trends and seasonal variations. The first one that is model 28 specifically deals with trends. And in the second module, we will discuss other seasonal variations.

(Refer Slide Time: 00:55)



Now, time trends I already have introduced once while discussing something else. No, this is a very common thing and what it measures is probably pretty easy to understand, but nevertheless, our basic discussion is here that much economic time series have a tendency to grow over a period of time, this could be because of slowly evolving preferences, technologies, demographics etcetera, this is known as trend or time trend.

So, when something is growing irrespective of the impact of other variables or factors, and this is because probably of a time factor then we can model it by using a time trend. While trying to draw causal inference using time series data, ignoring the fact that two sequences are trending in the same or opposite directions can lead us to falsely conclude that changes in one variable are actually caused by changes in another variable.

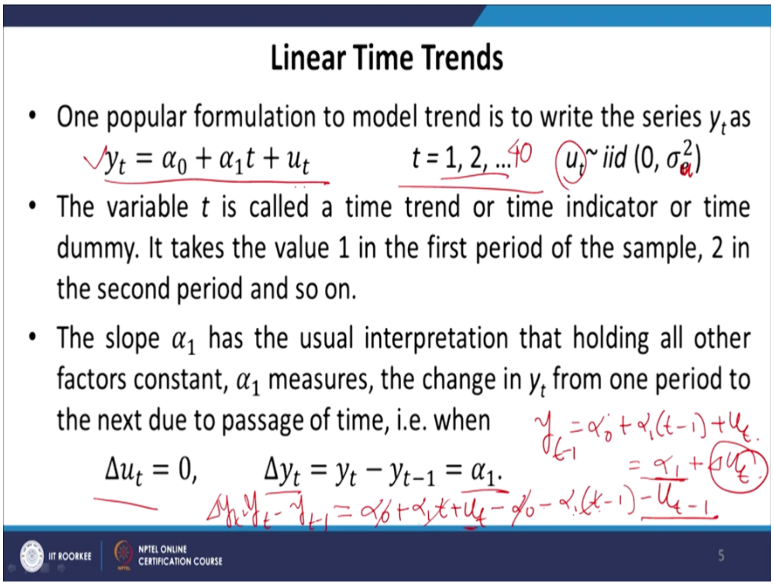
In many cases, the two-time series processes appear to be correlated only because they are both trending over time for reasons related to other unobserved factors. For example, by this, we simply want to mean that this is possible that there is a y series and there is an x series, one of them is increasing, the other one is decreasing monotonically.

Alternatively, it is also possible that both y and x are growing over a period of time. Now, we can observe or expect some correlation between them and the correlation would be obvious because when one is growing, the other one is also growing, but it is quite possible that they are completely unrelated to each other. And even if we observe a high correlation, it is not possible that they are correlated.

Here, we will focus on deterministic trends. Deterministic trends are trends that evolve in a perfectly predictable way. So far, we have talked about random variables, the opposite of something being random is being deterministic. So, which can be determined with 100 percent accuracy is called deterministic. So, when you talk about deterministic trend, then we certainly want to imply that this trend does not involve any stochastic component

If we are incorporating a deterministic trend, then it implies that this trend is constant over a period of time, the trend is having an impact on the variable if the trend is found to be significant, then it is having an impact on the variable we are concerned with, with 100 percent accuracy, so that trend part we can predict with 100 percent accuracy. So, if a variable is observed to grow over a period of time and if the inclusion of a trend explains nearly 99 percent of its variations, then it implies that we can predict the variable with almost 99 percent accuracy, just by making a dependent on time or the forecast are much easier. So today, the value of that series is 100. And we have observed that it has been growing at a rate of 1 percent over a period of time. So, we could say that in the next period it is going to be 101. So that prediction is there with a lot of determinism and that is why we call it a deterministic trend.

(Refer Slide Time: 04:39)



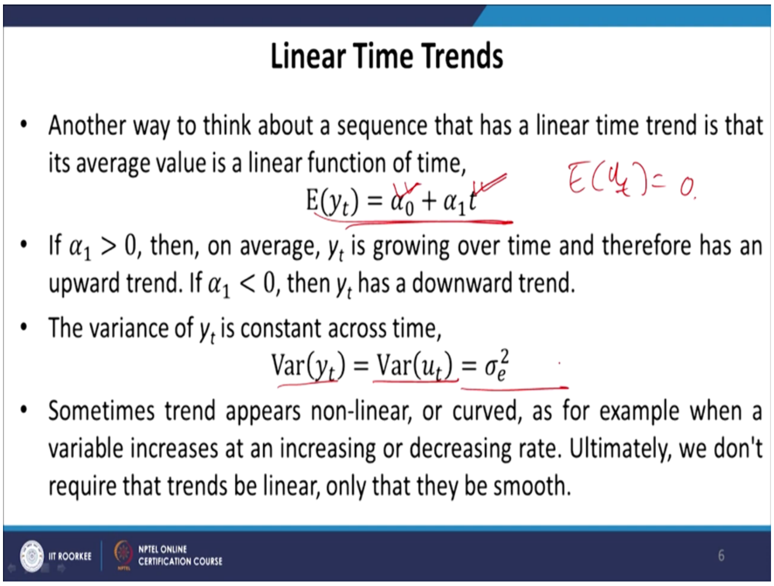
One popular formulation to model trend is to write the series (refer slide time: 4:45- 5:06), where, our error term I assume it to be well behaved. So, it is again a white noise process it is identically and independently distributed with 0 mean and constant variance at σ_e^2 or it can be

σ_u^2, and this is identically an independently distributed itself implies that there is no auto covariances or autocorrelations between the values of the error term.

The variable t is called a time trend or time indicator or time dummy, it takes the value one in the first period of the sample, two in the second period, and so on. The slope α_1 has the usual interpretation that holding all other factors constant, $\alpha_1$measures the change in y_t from one period to the next due to the passage of time. Now, in this expression, we do not have any independent variable.

So, this is we are focusing only on time and considering our time trend in as the factor responsible for explaining variations in y. Now, if I can write (refer slide time: 16:17- 7:54).

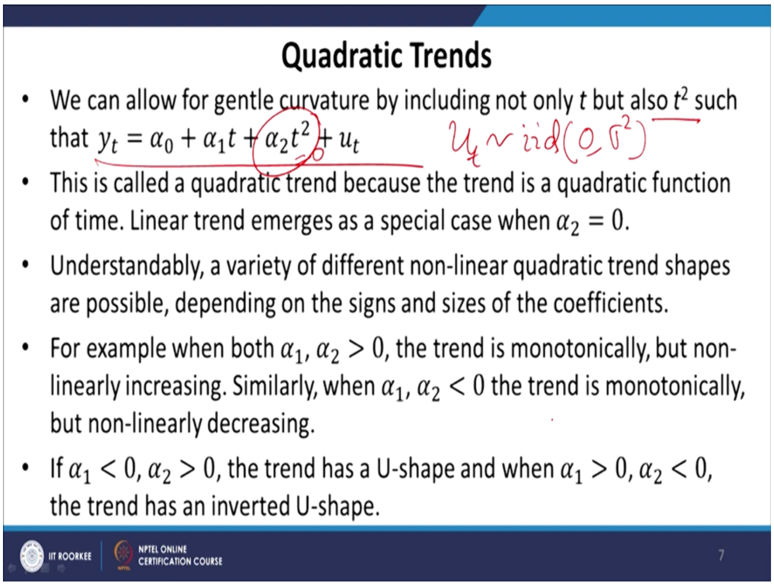
(Refer Slide Time: 07:53)



Another way to think about a sequence that has a linear time trend is that the average value is a linear function of time. So, if we take the expected value of this expression, then we would arrive at this (refer slide time: 8:07- 9:06).

Sometimes trend appears nonlinear or curved, for example, when a variable increases at an increasing or decreasing rate, then we cannot use a linear trend. So, the simplest form and the most popular form is the linear trend, but trends need not be always linear. So, in that case, we do not require the trends to be linear, what we want is that the trends to be smooth, and as a result of which we can also consider nonlinear trends.

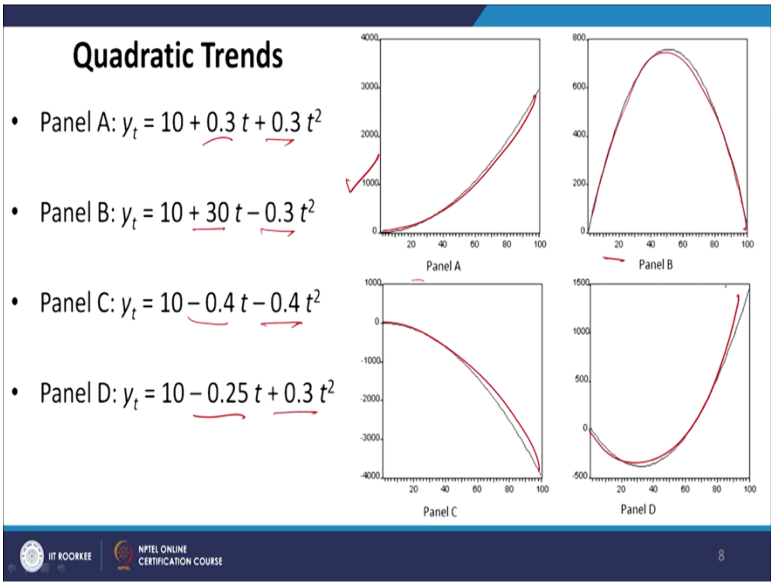
(Refer Slide Time: 09:40)



So, one kind of nonlinear trend is quadratic trends, we can allow for gentle curvature by including not only t, but also t^2, such that my original equation which was (refer slide time: 9:56- 10:40). Understandably, a variety of different nonlinear quadratic trend shapes are possible, depending on the signs and sizes of the coefficients.

For example, when both $\alpha_1, \alpha_2 > 0$, that trend is monotonically, but non-linearly increasing. So, I will be having pictures or graphs of monotonically but non-linearly increasing trends, quadratic trends. Similarly, when α_1, $\alpha_2 < 0$, the trend is monotonically, but non-linearly decreasing, if $\alpha_1 < 0$, $\alpha_2 > 0$that trend has a U-shape curve. And when $\alpha_1 > 0$ and $\alpha_2 < 0$, that trend has an inverted U-shape.

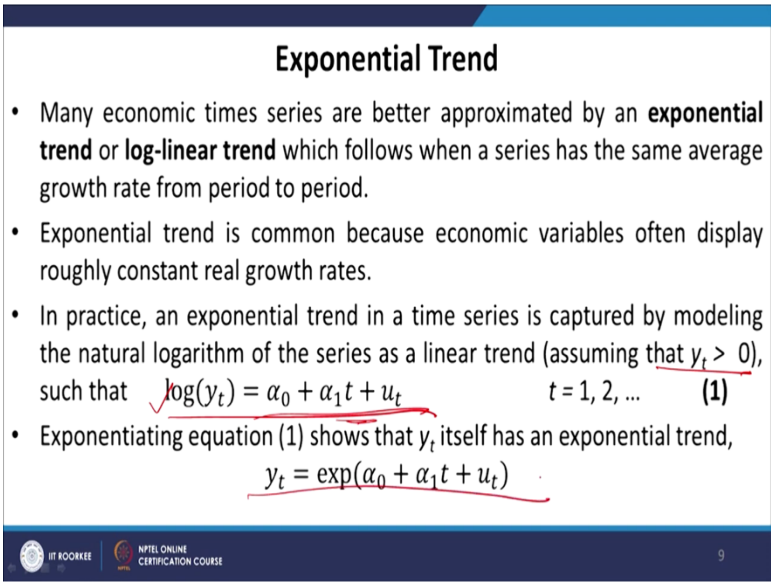
(Refer Slide Time: 11:35)



So, this is what we are showing here with examples, that in panel A, where we have both coefficients positive, so we have a monotonically increasing quadratic trend. In panel B, we have one positive one negative, so this is panel B. So, this is an inverted U-shaped trend. Panel C has both negative coefficients and that is how we have a monotonically decreasing trend.

And finally, when we have a negative coefficient for t and a positive coefficient for the quadratic term t^2, then we have a U-shaped time trend. So, the quadratic trend is more useful from the perspective that if we plot a series and plot a series against time, and we do not see it to be much linear, then quadratic plots can be or quadratic trends can be of much help that will give us a better fit of the model. Without considering any explanatory variable a quadratic trend may explain a major part of the variations in our dependent variable.

(Refer Slide Time: 12:48)



Another type of nonlinear trend is an exponential trend, but this is again a very popular and common trend that we often use in economics and this is popular also because this can be converted into a linear form. So, many economic time series are better approximated by an exponential trend or log-linear trend, which follows when a series has the same average growth rate from period to period.

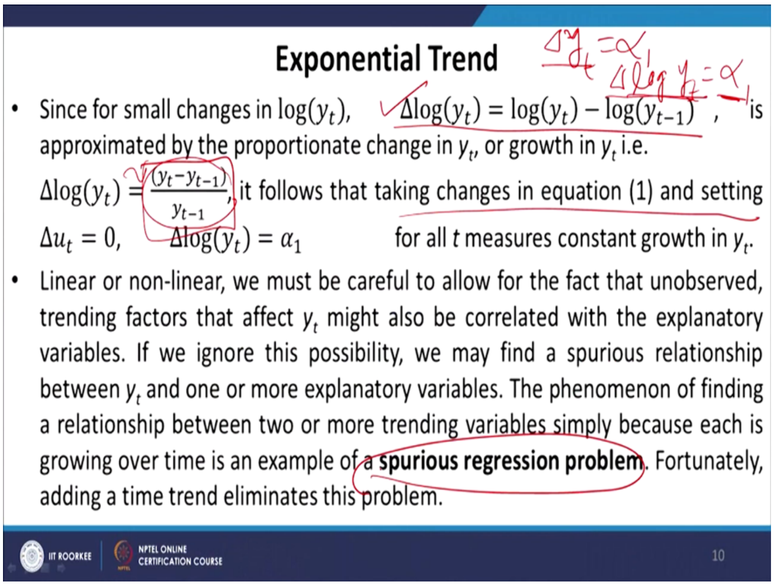
The exponential trend is common because economic variables often display roughly constant real growth rates. So, if we observe that the real GDP of an economy has been growing for the last 10 years at a rate of 2 percent per annum, then this can be probably plotted as an exponential trend. In practice, an exponential trend in a time series is captured by modelling in the natural logarithm of the series as a linear trend, and in order to consider natural logarithm, we must have $y_t > 0$ as you know that we cannot have a logarithm of negative values.

So, provided that the series takes only positive values we can go for a natural logarithm, so in case most often, if we are working already with growth rates, for example, inflation. Inflation can be positive as well as negative. So, in that case, we may not consider an exponential trend.

But for most other series, for example, GDP in its absolute values, either real or nominal can be always considered for an exponential trend model. So, if $y_t > 0$, we can go for an exponential trend and it is written as (refer slide time: 14:38). You can see that the right-hand term remains the same as that of a linear trend.

We call it a log-linear trend because or log-linear model, because we have logarithm on the left-hand side, and the right-hand side is in a linear form, and it is a log-linear trend because this model trying to incorporate our trend term. Exponentiating equation 1 shows that y_t itself has an exponential trend. So, (refer slide time: 15:11).

(Refer Slide Time: 15:16)



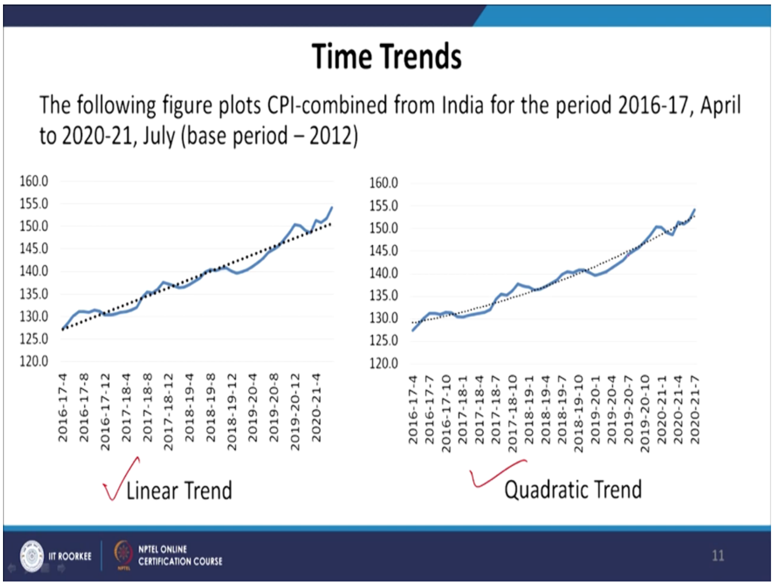
Since, for small changes, we know that (refer slide time:15:20- 16:40). So, here if we are going for an exponential trend, then α_1 measure the constant change in the dependent variable, constant growth in the dependent variable. So, what it says is that with time the series is growing at a constant rate. And what is that constant rate, the rate is α_1.

Linear or nonlinear, we must be careful to allow for the fact that unobserved trending factors that affect y_t might also be correlated with explanatory variables. If we ignore this possibility, we may find a spurious relationship between y_t and one or more explanatory variables. This phenomenon of finding a relationship between two or more training variables simply because each is growing over time is an example of a spurious regression problem.

Fortunately, adding a time trend eliminates this problem. So, it is possible that independent variables and dependent variables both are functions of time trend, both have some components or both are growing just because they are growing with time and they are not related with each other. This problem has also been probably discussed at the beginning of this module, that it is possible that they are unrelated and we observe relationships between them, just because both of them are growing over a period of time.

Now, if they are unrelated, but we observe a relationship between them by running a regression, then we call it a spurious regression. And, this problem can be considered or tackled by simply including a time trend into the regression analysis, that will take care of any trend component that might be impacting the independent as well as the dependent variables.

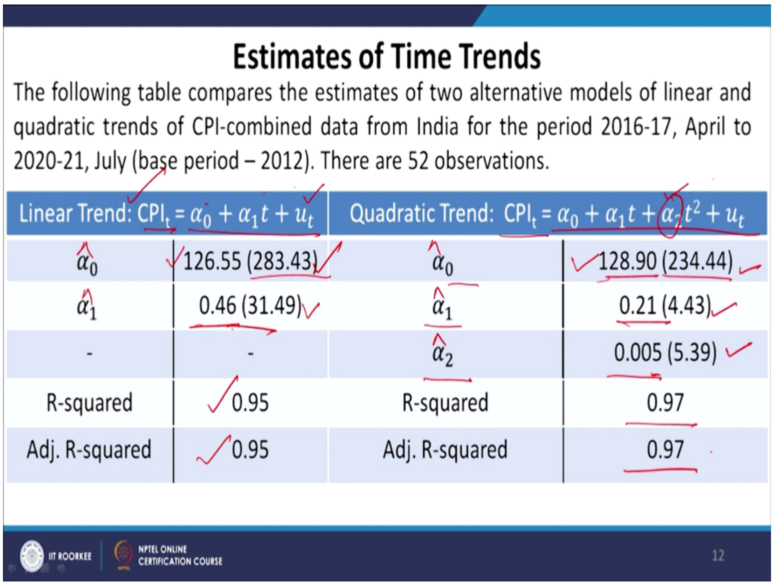
(Refer Slide Time: 18:32)



So, we now take an example, the following figure plots CPI-combined from India for the period 2016-17, April. These are monthly figures to 2020-21, July with the base period 2012. So, we have plotted one linear trend and one quadratic trend, these are simply generated by taking the data into Excel and the graphs are plotted the trend lines are added. You can see that the linear trend is quite a good fit.

And when you go for quadratic trend, the curvature is very small or minute, one can also ignore such a curvature, the linear trend in itself is quite a good fit.

(Refer Slide Time: 19:20)



Estimates of Time Trends

The following table compares the estimates of two alternative models of linear and quadratic trends of CPI-combined data from India for the period 2016-17, April to 2020-21, July (base period – 2012). There are 52 observations.

Linear Trend: $CPI_t = \alpha_0 + \alpha_1 t + u_t$		Quadratic Trend: $CPI_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + u_t$	
$\hat{\alpha}_0$	126.55 (283.43)	$\hat{\alpha}_0$	128.90 (234.44)
$\hat{\alpha}_1$	0.46 (31.49)	$\hat{\alpha}_1$	0.21 (4.43)
-	-	$\hat{\alpha}_2$	0.005 (5.39)
R-squared	0.95	R-squared	0.97
Adj. R-squared	0.95	Adj. R-squared	0.97

So, in a go for estimation of the series, what we do is first of all for linear trend, we have (refer slide time: 19:28- 19:45).

So, the following table compares the estimates of two alternative models of linear and quadratic trends of CPI combined data from India for the period 2016-17, April to 2020-21, July with the base period 2012 and there are 52 observations. So, there are 52 observations, you can understand that the degrees of freedom for this model would be 50 and the degrees of freedom for this model would be specifically the residual degrees of freedom for this is 50 and the residual degrees of freedom for this model is 49.

This is 52 minus 3, this is 52 minus 2 there are only two variables. Now, just looking at the estimates we see that (refer slide time: 20:29).

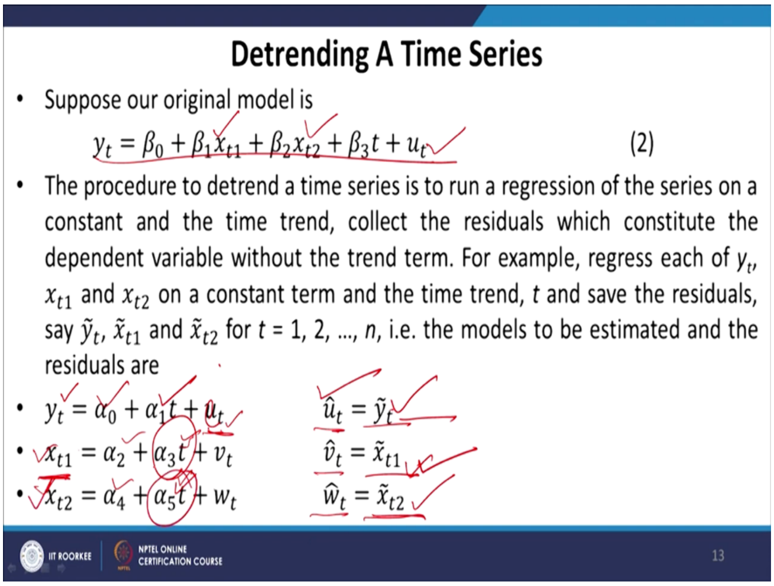
The t-statistics are measured in the parenthesis. But what we observe is that, when we just have a linear trend, the linear trend explains 0.46 percent of the variations in the dependent variable that is the CPI combined series. So, what it says is that, as time changes or with time, when you move from one month to the next month, CPI increases by 0.46.

The R-squared value is 95 percent, adjusted R-squared is also 95 percent, which implies that 95 percent of the variations in the dependent variable are explained. And when we consider the quadratic trend, what we observe is that α_0 is again a big value, α_1 has likely gone come down now, but this is still significant, it is now 0.21.

And α_2 the coefficient of the quadratic term is very small, but it is also significant, it is 0.005. I have mentioned here only up to three decimal points because otherwise for two decimal points it would have been simply 00. So, the quadratic term though significant statistically is not much perceptible, the impact is, or probably the shape or the contribution of the quadratic term is very small to the dependent variable.

R-squared value, as well as adjusted R-squared value, has also increased, the inclusion of the quadratic terms since it is significant explains 97 percent of the variations in the dependent variable that is CPI-combined.

(Refer Slide Time: 22:42)



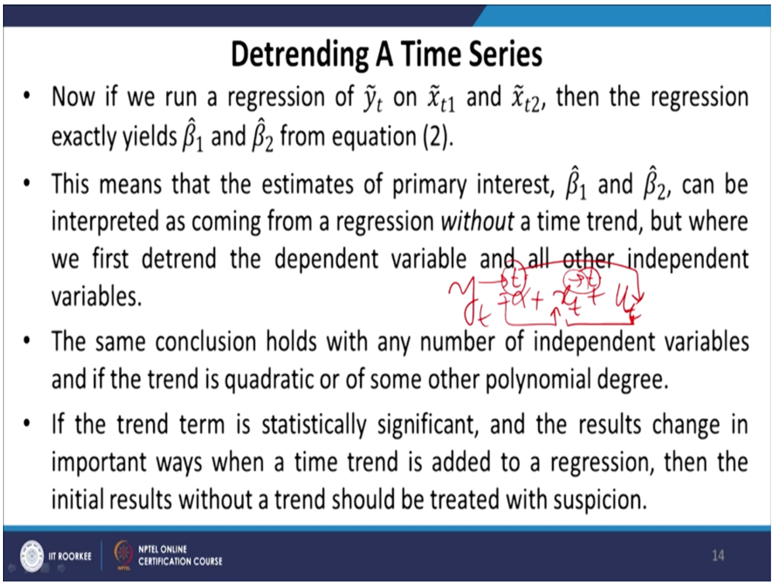
Now, we have talked about detrending a time series, how we can remove the trend component of a time series. Either we can incorporate trends, or we can also work with detrended series. Now, this is not important, in this sense, we do not need to detrend a model or detrend a series, but what it shows is that, if we go for the detrending a series then if we work with detrended series, then we are going to get back the exact same estimators if you would have included a trend in the original model.

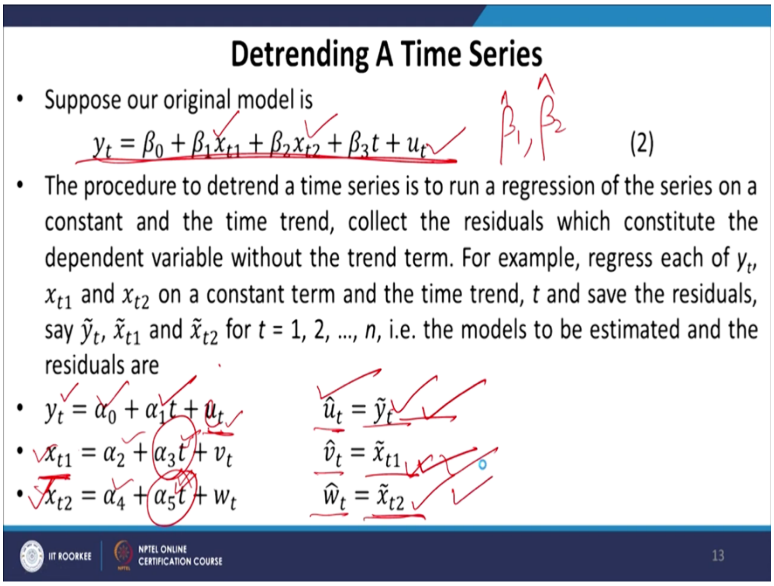
So, suppose this is my original model where (refer slide time: 23:28). So, I have considered two independent variables, there can be k independent variables. The procedure to detrend a time series is to run a regression of the series on a constant and the time trend. So, if you run a regression on the constant and the time trend, collect the residuals which constitute the dependent variable without the trends term.

For example, regress each of (refer slide time: 24:02- 24:53), because in the next step, what I am going to do is that regress this residual on these independent variables. Now, the point is that if these variables do not have any trend term or any variable does not have a trend term, then the trend term will be insignificant.

And this residual and this one will not differ much, but in case there is a trend term, this trending term will be eliminated. Now, $\tilde{y}_t$ is a series that is the original series except for the impact of the trend factor. Similarly, $\tilde{x}_{t1}$ is the original series without the impact of the trend factor and $\tilde{x}_{t2}$ is again the original series without the impact of the trend term or trend factor.

(Refer Slide Time: 25:46)





Now, if we run a regression of (refer slide time: 25:50).

This means that the estimates of primary interest $\hat{\beta}_1$ and $\hat{\beta}_2$ can be interpreted as coming from a regression without a time trend, but where we first detrend the dependent variable and all other independent variables. The same conclusion holds with any number of independent variables and if the trend is quadratic, or of some other polynomial degree. So, in a similar fashion, we can detrend any series having a quadratic trend or a trend of any other polynomial series.

If the trend term is statistically significant, the results change in important ways by the time trend is added to a regression, then the initial results without a trend should be treated with suspicion, which implies that if the trend is significant, then it must be included into the regression analysis.

Otherwise, my regression results might not be valid, because, simply because of the reason that an important variable has been dropped from the equation. Also, it is possible that I am estimating an equation like (refer slide time: 27:24- 27:51). And as a result of which my CLRM assumptions will be violated. The regression results do not fulfill the Gauss-Markov or CLRM assumptions, my results are not trustworthy, they cannot be dependent upon, they are not reliable results. As a result of which we are saying that if the trend term is significant, they must be included in the regression analysis.

So that is all about the inclusion of trends in a regression analysis, the possible forms, their impact, and how we can go for the detrending, in the next module on-trend and modelling seasonality. I will be dealing with other types of seasonality or other seasonal variations that we experience in the time series, or that economic and business time series. Thank you.