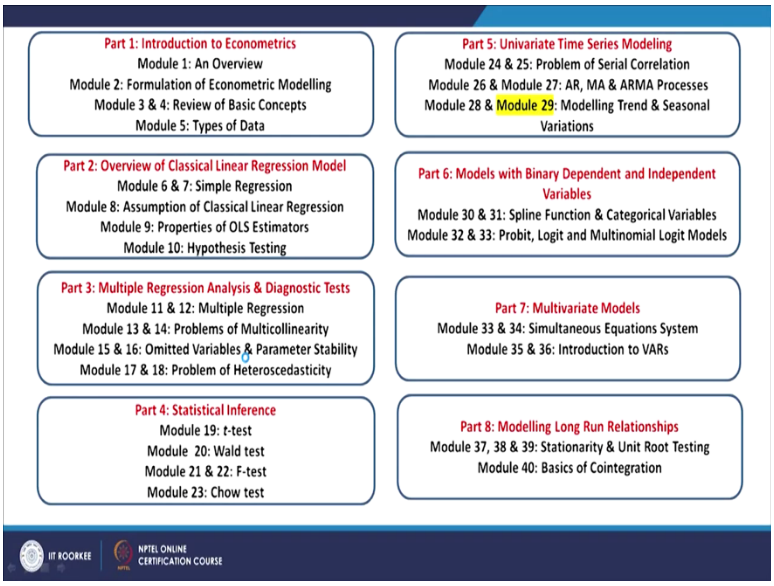
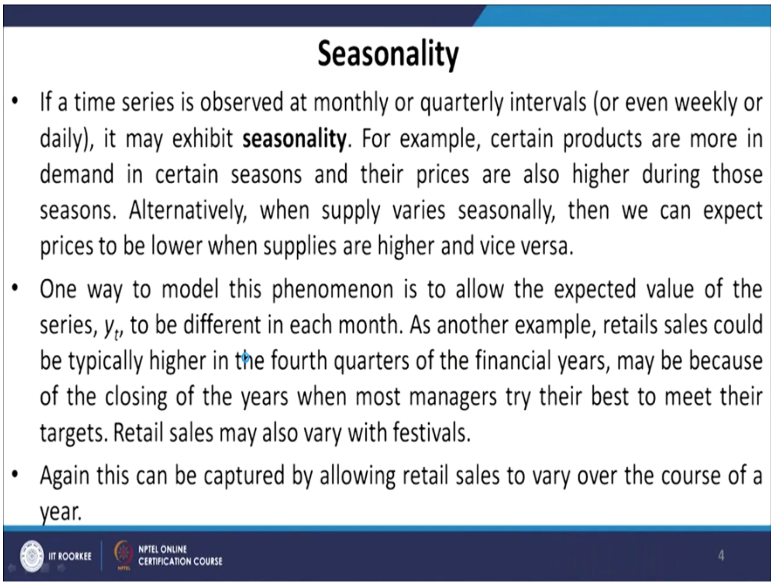
Econometric Modelling
Professor Sujata Kar
Department of Management Studied
Indian Institute of Technology Roorkee
Lecture 29
Modelling Trend & Seasonal Variations - II

(Refer Slide Time: 00:37)



Hello everyone, this is module 29 of the course on econometric modelling, we have been discussing the univariate time series model. And under that in the last module, I had discussed modelling trends. And in this module, I am going to discuss seasonal variations or modelling seasonal variations or seasonality. So, now focusing on seasonal variations.

(Refer Slide Time: 00:50)



First of all, we would be defining seasonality. If a time series is observed at monthly or quarterly intervals, or even weekly or daily, it may exhibit seasonality. So, whenever there are very explicit clear-cut variations in the data observed, just because of a change in the frequency of observation, then that can be ascribed to the seasonality component. For example, certain products are more in demand in certain seasons and their prices are also higher during those seasons.

Alternatively, when supply varies seasonally, then we can expect prices to be lower when supplies are higher and vice versa. So, even if we do not talk about prices, we can understand that for certain products, the demands are higher in certain seasons. For example, during the summer months, we would find that the demand for AC, fans, coolers etcetera will be higher.

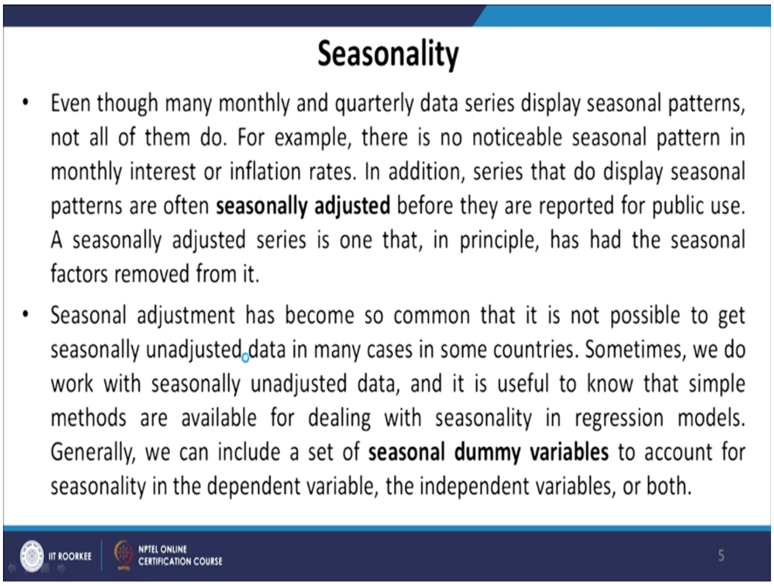
Similarly, during the winter season, we must be observing demand for room heaters to be higher and other appliances which provide us heating services, maybe even electric kettle, the products which are used for either heating air or water. So, as a result of which we can observe seasonal variations so these variations which we are observing in the demand for a particular product is not due to a change in income, change in prices, or something it is just because that time of the year has changed.

So, the season requires a different kind of appliance. And that is why we ascribe it to seasonality, seasonal variations. So, one way to model this phenomenon is to allow the expected value of this series y_t to be different in each month if we consider monthly data or each quarter if we consider quarterly data and so on.

As another example, retail sales could be typically higher in the fourth quarter of the financial years, by financial year, I mean the year starting from April and ending on 31st of March every year. So, the last quarter is January, February, March, we can expect a higher retail sales during the last quarter, maybe because of the closing of the years when most managers try their best to meet their targets.

Our retail sales may also vary with festivals, for example, it is very pronounced in the western economies that during Christmas the retail sales increases there. And similarly, in our country, we can expect an increase in the purchase or demand for jewelry, gold products, as well as garments during certain seasons like when marriages take place and then also during the festivals like Diwali, Holi we can expect an increase in the demand for certain goods of consumption. Again, this can be captured by allowing retail sales to value over the course of a year. So, this is how we generally could incorporate a seasonal variation.

(Refer Slide Time: 04:15)



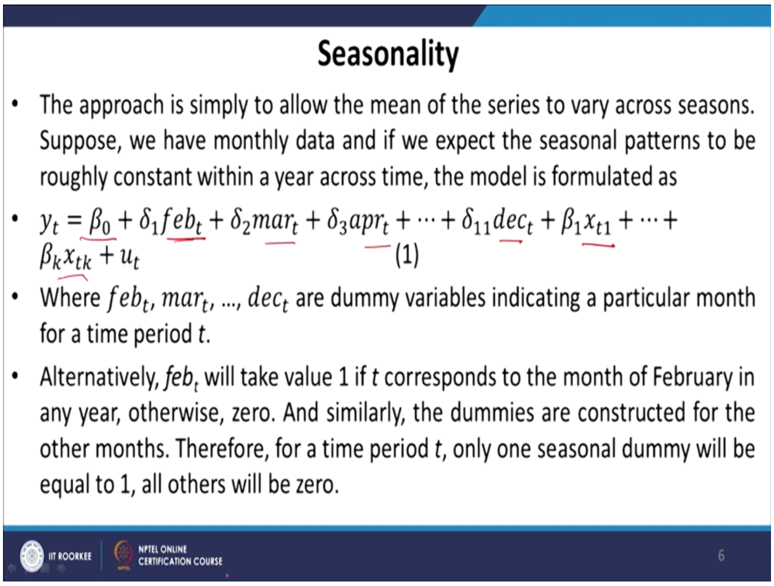
Even though many monthly and quarterly data series display seasonal patterns, not all of them do. For example, there is no noticeable seasonal pattern in monthly interest or inflation rates. In addition, series that do display seasonal patterns are often seasonally adjusted before they are reported to public use. A seasonally adjusted series is one that in principle has had the seasonal factors removed from it and we work with a seasonally adjusted series.

So, some processes or methods, some rudimentary or basic processes of deseasonalizing data will be discussed in this module. But then there are other programs also available which automatically deseasonalize a series or a set of data. The seasonal adjustment has become so common that it is not possible to get seasonally unadjusted data in many cases in some of the countries specifically if we can talk about US data, sometimes we do work with seasonally unadjusted data as well.

And it is useful to know that simple method are available for dealing with seasonality in the regression model. Also, another problem is that having a seasonal component, a seasonal component explains certain variations in the dependent variable. So, removing the seasonal component is losing out on some information.

But of course, we know very well that the data exhibit some seasonal patterns and if you are not interested in understanding estimating the seasonal patterns, then this is another variation that can be removed, and then we work with the seasonally adjusted series where we are more interested in examining the impact of other factors on that particular variable. So, generally, we can include a set of seasonal dummy variables to account for seasonality in the dependent variable, the independent variables, or both.

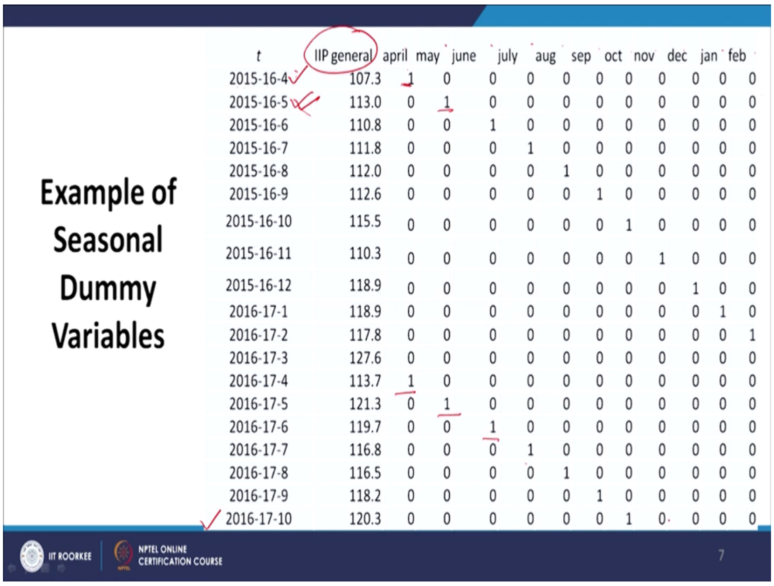
(Refer Slide Time: 06:15)



The approach is simply to allow the mean of the series to vary across seasons. Suppose, we have monthly data, and if we expect the seasonal patterns to be roughly constant within a year across time, then their model is formulated as having 11 dummy variables for 11 seasons like February, March, April, May, June, July, August, September, October, November, and December.

Other than that, I am also including some of the independent variables (refer slide time: 6:48-7:37)

(Refer Slide Time: 07:37)



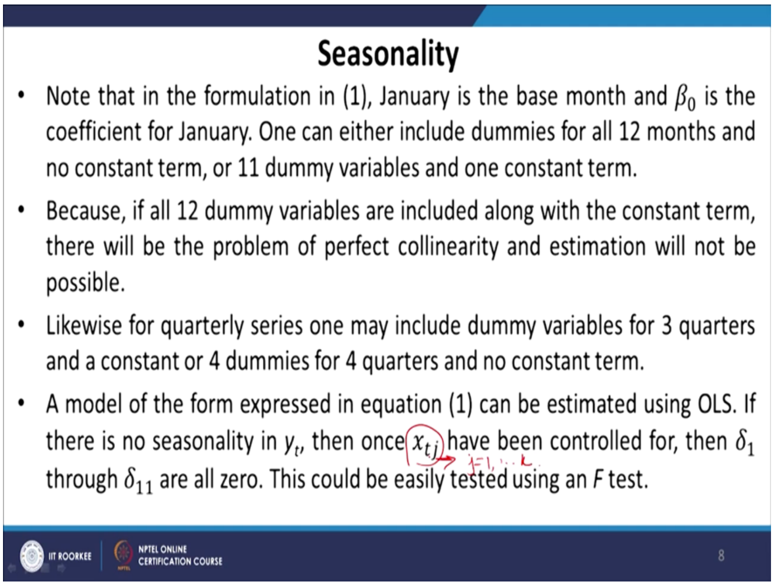
t	IIP general	april	may	june	july	aug	sep	oct	nov	dec	jan	feb
2015-16-4	107.3	1	0	0	0	0	0	0	0	0	0	0
2015-16-5	113.0	0	1	0	0	0	0	0	0	0	0	0
2015-16-6	110.8	0	0	1	0	0	0	0	0	0	0	0
2015-16-7	111.8	0	0	0	1	0	0	0	0	0	0	0
2015-16-8	112.0	0	0	0	0	1	0	0	0	0	0	0
2015-16-9	112.6	0	0	0	0	0	1	0	0	0	0	0
2015-16-10	115.5	0	0	0	0	0	0	1	0	0	0	0
2015-16-11	110.3	0	0	0	0	0	0	0	1	0	0	0
2015-16-12	118.9	0	0	0	0	0	0	0	0	1	0	0
2016-17-1	118.9	0	0	0	0	0	0	0	0	0	1	0
2016-17-2	117.8	0	0	0	0	0	0	0	0	0	0	1
2016-17-3	127.6	0	0	0	0	0	0	0	0	0	0	0
2016-17-4	113.7	1	0	0	0	0	0	0	0	0	0	0
2016-17-5	121.3	0	1	0	0	0	0	0	0	0	0	0
2016-17-6	119.7	0	0	1	0	0	0	0	0	0	0	0
2016-17-7	116.8	0	0	0	1	0	0	0	0	0	0	0
2016-17-8	116.5	0	0	0	0	1	0	0	0	0	0	0
2016-17-9	118.2	0	0	0	0	0	1	0	0	0	0	0
2016-17-10	120.3	0	0	0	0	0	0	1	0	0	0	0

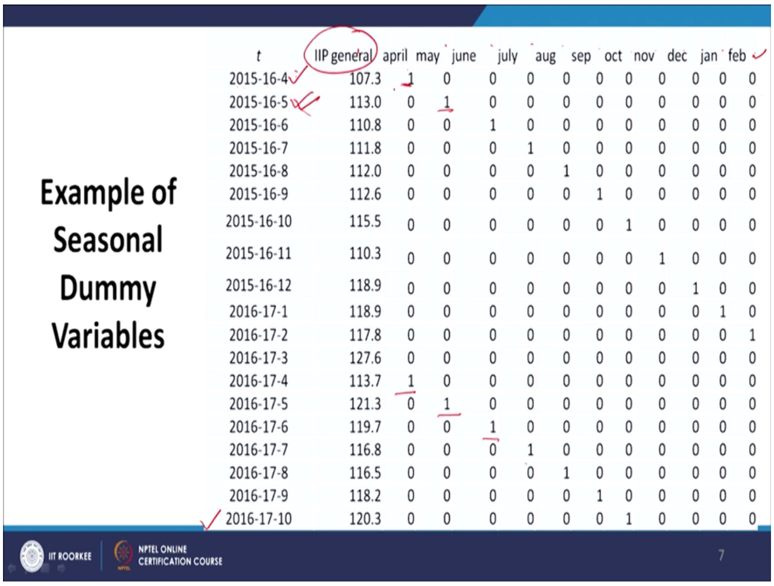
So, there is an example of seasonal dummy variables, what I have considered here is IIP general, I have taken the data from India, the data is long after 2020-21, July, but I have truncated it here because I just want to show here, how are the data seasonal dummy variables are included. So, here are the first two-component 2015-16 refers to the financial year and the last component refers to the month. So, 4 here refers to April, 5 May, and so on.

So, you can see that I have not included any other independent variables also, and the constant term need not be included, because most packages automatically report the parameter estimate of a constant term, if only we do not want the estimate of a parameter constant term or if we want to exclude the constant term, we can specify that in that regression package and accordingly that will be removed. But otherwise, this is the default setup.

So, we are having 1 to 11 dummy variables starting from April, 4 is the month of April. So, April is having one here for the rest of the months I am having 0. Similarly, this is May, so again for May, I have one for the rest of the month it is 0. And you can see that the same thing is being repeated here. Again, one for April, one for May, one for June, and so on.

(Refer Slide Time: 09:15)





t	IIP general	april	may	june	july	aug	sep	oct	nov	dec	jan	feb
2015-16-4	107.3	1	0	0	0	0	0	0	0	0	0	0
2015-16-5	113.0	0	1	0	0	0	0	0	0	0	0	0
2015-16-6	110.8	0	0	1	0	0	0	0	0	0	0	0
2015-16-7	111.8	0	0	0	1	0	0	0	0	0	0	0
2015-16-8	112.0	0	0	0	0	1	0	0	0	0	0	0
2015-16-9	112.6	0	0	0	0	0	1	0	0	0	0	0
2015-16-10	115.5	0	0	0	0	0	0	1	0	0	0	0
2015-16-11	110.3	0	0	0	0	0	0	0	1	0	0	0
2015-16-12	118.9	0	0	0	0	0	0	0	0	1	0	0
2016-17-1	118.9	0	0	0	0	0	0	0	0	0	1	0
2016-17-2	117.8	0	0	0	0	0	0	0	0	0	0	1
2016-17-3	127.6	0	0	0	0	0	0	0	0	0	0	0
2016-17-4	113.7	1	0	0	0	0	0	0	0	0	0	0
2016-17-5	121.3	0	1	0	0	0	0	0	0	0	0	0
2016-17-6	119.7	0	0	1	0	0	0	0	0	0	0	0
2016-17-7	116.8	0	0	0	1	0	0	0	0	0	0	0
2016-17-8	116.5	0	0	0	0	1	0	0	0	0	0	0
2016-17-9	118.2	0	0	0	0	0	1	0	0	0	0	0
2016-17-10	120.3	0	0	0	0	0	0	1	0	0	0	0

Note that in the formulation in equation 1, January is the base month, and β_0 is the coefficient for January. So, since we are considering 11 dummy variables and there are 12 months, so we are not considering the 12 months, instead, there is a constant term and the constant term corresponds to the coefficient of the dummy which is not included or mentioned explicitly.

One can either include dummies for all 12 months and no constant term or 11 dummy variables and one constant term which we have just done because if all dummy variables are included, along with the constant term, there will be the problem of perfect collinearity and estimates will not be possible.

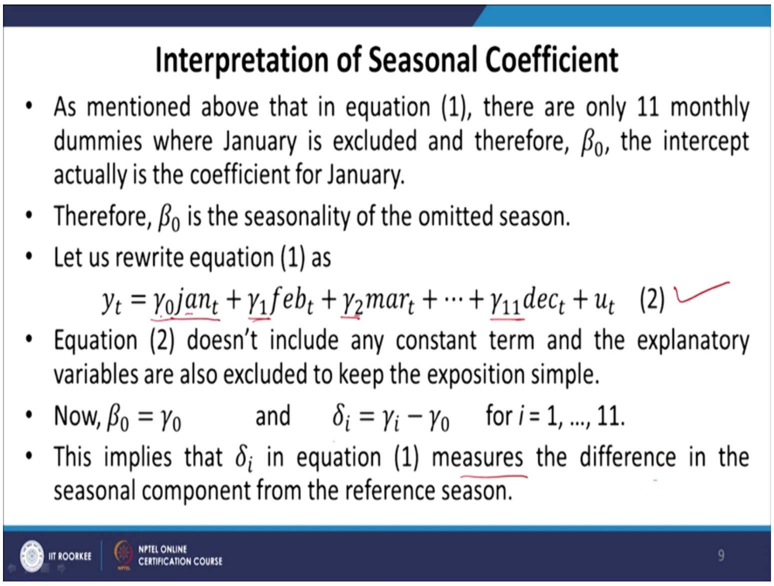
As we can understand that if we have all 12 dummy variables here, for example, if I have also the month of March either here or here, and if I sum all these 12 columns, then that will also be equal to 1, we already have a constant term having all observations equal to 1, then the sum of these 12 dummies will also have all observations equal to 1.

So, there will be the problem of perfect multicollinearity and as a result of which estimates will not be generated, because estimates are not unique these problems have already been discussed in a previous module. So, likewise for quarterly series, one may include dummy variables for 3 quarters and a constant term or 4 dummies for 4 quarters and no constant term. A model of the form expressed in equation 1 can be estimated using OLS.

If there is no seasonality in (refer slide time: 11:05). They would all be 0, if there is no seasonality, this could be easily tested using an F-test. So, we can go for a restricted regression and an unrestricted regression.

So, where the restricted regression does not include any dummy variable, seasonal dummy variables and the unrestricted regression does include the original model as expressed in equation 1. Now, (refer slide time: 11:48).

(Refer Slide Time: 12:09)



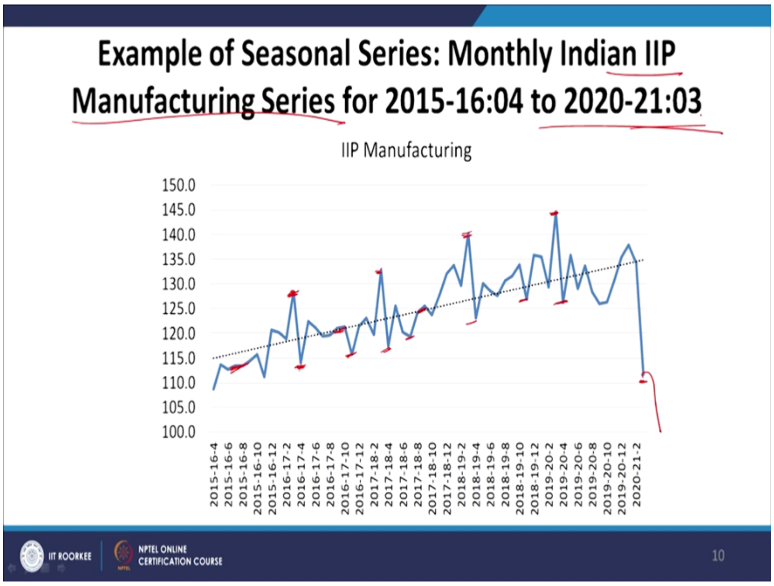
As mentioned in equation 1, there are only 11 monthly dummies, where January is excluded and therefore, β_0, the intercept is the coefficient for January. Therefore, β_0 is the seasonality of the omitted season. So, if we rewrite equation 1 by including the January dummy as well. So, now, instead of having a constant term I am having all the 12 dummies, then equation 2 does not include any constant term.

And the explanatory variables are also excluded just to keep the exposition simple because even if I exclude the independent variables, the following discussion is not going to be any different. So, now, you can observe that β_0, the original intercept is now equal to γ_0 and $\delta_i = \gamma_i - \gamma_0$, where $i = 1, \ldots. 11$.

So, (refer slide time: 13:24). So, now, we are trying to get into the interpretation of the coefficients of the seasonal dummy variables.

What it says is that the coefficients of the seasonal dummy variables which are there when we do not have all 12 dummies, but rather when we have 11 dummy variables and one intercept, then the coefficients of those dummy variables measure the difference in the seasonal component from the reference period. For example, here the reference period is January.

(Refer Slide Time: 14:18)



So, now I take up an example of seasonal series again working with monthly IIP manufacturing series, so 2015-16, April; 2020-21, March. In the Indian context, most of the data are reported for financial years, that is from April to March. And that is why, though the original equation had calendar months like January, February, and so on. In this example, I am considering periods like starting from April.

So, as you can see that this graph very explicitly shows that there is seasonal variation in IIP manufacturing. These peaks correspond to March of every financial year. So, this is like 2016-17, March, this is 17-18, March, this is 18-19, March, and so on. So, in each and every year, the IIP manufacturing, Indian IIP manufacturing series reaches its peak and then after March there is a sharp drop every year. So, there is a very clear seasonal pattern.

And then after April, again there is an upward movement there is a downward movement most often you would see that there are ups and downs on a continuous basis. Now, there are certain seasons where not many variations are observed. For the most part, it is possible that not many variations are observed during certain seasons, but the prominent differences are what pronouncedly come out is that in March it reaches a peak, and across all the years, the month of March records the maximum increase or the highest figure for Indian IIP manufacturing series.

And similarly, April almost corresponds to the lowest values of the series, but then there are also other months when it is somewhat equal to April, but March is not matched by any other month in terms of the values of the series. We also observe that there is a broadly upward trend. Now, this broad upward trend implies that there is also a trend component.

Another thing I just would like to mention is that I have considered data till 2020-21, March because after that, there was a sharp drop in the IIP manufacturing and it was very low. As you can see that it went up to this level and it was further down, it came below 100 for 2 months during 2020-21 most probably in July-August. So, for one month, the figure was 53 the other was 79.

So, because of these outliers, that is extreme observations which are not getting well aligned with the rest of the series, the regression results including those observations, were very poor, but the moment I excluded those extreme values and truncated my data at 2020-21, March, I had a very nice regression result.

(Refer Slide Time: 17:47)

Estimates of IIP Manufacturing

Reference season: April

No. of observations=59

$R^2 = 0.66$

Adj. $R^2 = 0.57$

	Coefficients	Standard Error	t Stat	P-value
Intercept	109.9833	2.582279	42.59155	3.34E-39
may	7.324931	3.360881	2.179467	0.034342
june	3.849861	3.36161	1.145243	0.257907
july	3.874792	3.362825	1.152243	0.255048
aug	4.099722	3.364526	1.218514	0.22911
sep	4.284653	3.36671	1.272653	0.209401
oct	4.429583	3.369378	1.314659	0.195004
nov	2.334514	3.372529	0.692215	0.492208
dec	8.679444	3.376161	2.570803	0.013373
jan	9.384375	3.380272	2.776219	0.007872
feb	5.349306	3.384861	1.580362	0.120731
mar	10.19424	3.389926	3.007215	0.004224
t	0.315069	0.040422	7.794467	5.17E-10

IIT ROORKEE | NPTEL ONLINE CERTIFICATION COURSE 12

So, for these regression results, I had two alternative models and they reported here. First of all, I have considered the reference period March and then the variables are the coefficients and their standard errors t values and t-statistics are plotted here along with an estimate of a time trend deterministic trend, you can see that all the p values are very small, which implies that all these months are significant.

So, they are significantly explaining the variations in the original series that is IIP manufacturing. And since we observe that March always has the highest figures or values as a result of which that is a reference to the base period, all the months are having negative coefficients. All the coefficients are negative, because the reference to the base period, the values are all lower. And in fact, April has recorded the maximum negative number or the lowest value.

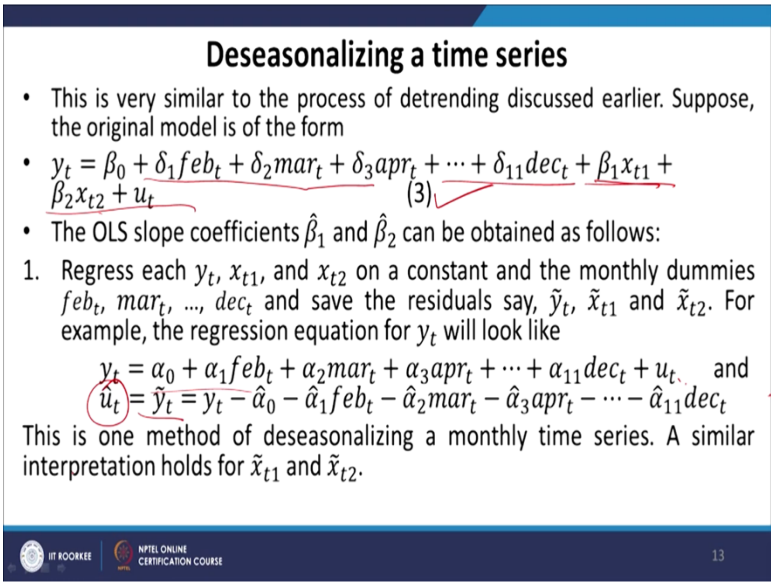
And the intercept, which measures the coefficient of the intercept term is very high. The t-statistic associated with it is also very high. The time trend or deterministic trend explains nearly 36 percent of the variations in the dependent variable holding other things constant. We also have observed a very good R-square value which is 91 percent and an adjusted R-squared value at 88 percent.

Now, I have re-estimated this model with April as the reference season. So, first was the season of March which reported the maximum figures for IIP. Now, April used to have the lowest figures for IIP. Now, you can see that against April, all the other months are having coefficients, positive coefficient, which implies that in reference to the base periods all the months are having positive contributions to the variations in the dependent variable.

And March is having the highest value among all other months, again the t-value explaining roughly 31 percent. You can see that against April there are two months which are insignificant, not two months rather there are quite a number of months like starting from June, July, August, September, October, November, all of them are having insignificant values, implies that with reference to the base period, their contributions to the dependent variable is not statistically significantly different.

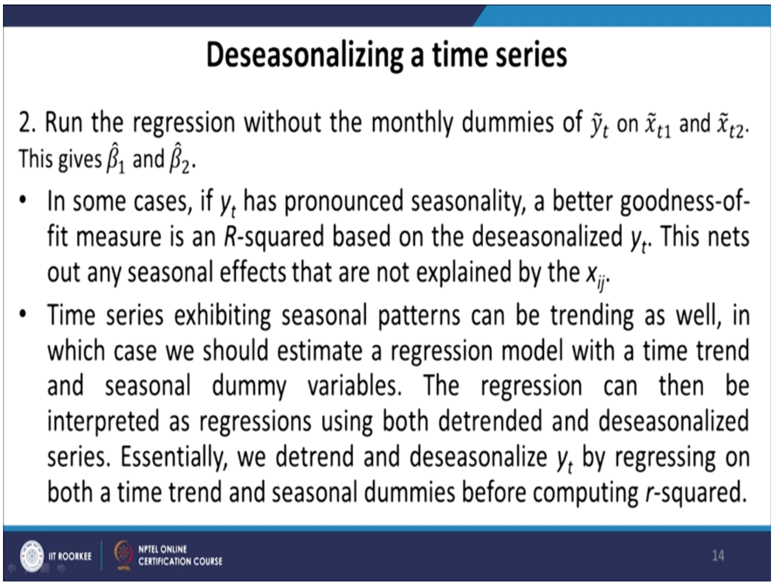
Here, the R-squared value is somewhat lower 66 percent, adjusted R-squared value is at 57 percent. So, this is how if we change the reference period or the base period, then we can observe differences in the coefficient estimates of the seasonal components. This is because the seasonal components measure its contribution to the dependent variable with reference to the base period. The difference between its contribution and the reference period is getting reflected here.

(Refer Slide Time: 21:11)



Now, we talk about how we can deseasonalize a time series. This is very similar to the process of detrending discussed earlier. Suppose, (refer slide time: 21:21- 22:34).

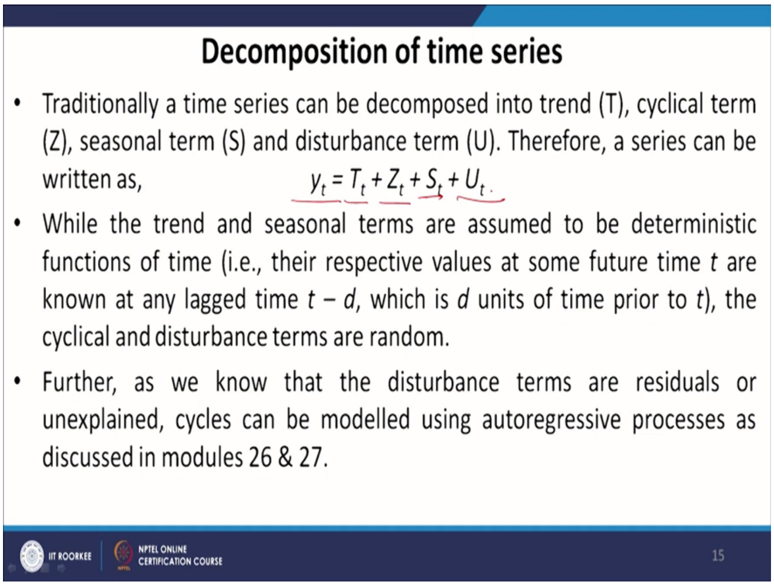
(Refer Slide Time: 22:33)



And after that we run the regression without the monthly dummies of (refer slide time: 22:39-23:08). Time series exhibiting seasonal patterns can be trending as well, in which case we should estimate a regression model with a time trend and seasonal dummy variables, the one that I have done in the example that I just showed you, the regression can then be interpreted as regressions using both detrended and deseasonalized series.

Essentially, we detrend and deseasonalized y_t by regressing on both a time trend and seasonal damages before computing R-squared.

(Refer Slide Time: 23:39)

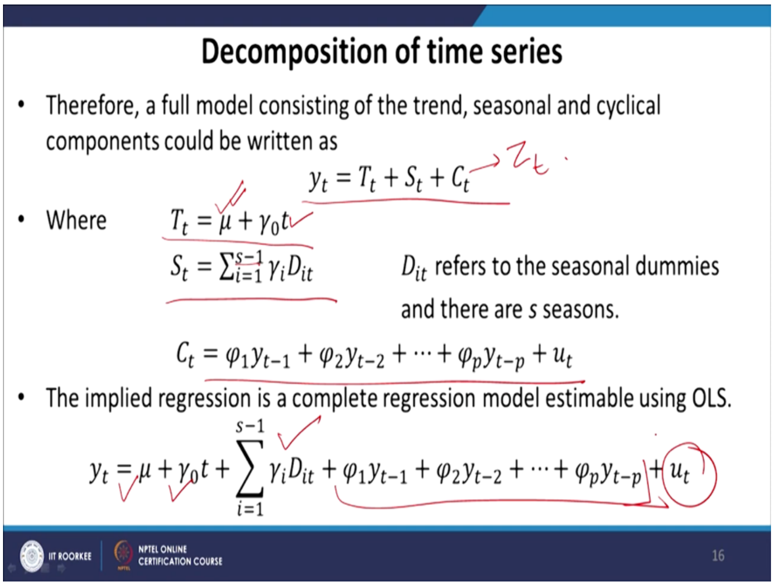


Now, we sum up whatever we have discussed in terms of all while discussing univariate time series modelling. So, traditionally a time series can be decomposed into a trend component, the deterministic trend upward or downward, cyclical terms, seasonal terms, and disturbance term. Therefore, a series can be written as (refer slide time: 24:03).

While the trend and seasonal terms are assumed to be deterministic functions of time, that is they can be predicted with 100 percent accuracy. That is their respective values at some future time t are known at any lagged time $t - d$ which is d units of time prior to t, the cyclical and disturbance terms are random terms. Further, as we know that the disturbance terms are residuals or unexplained, cycles can be modelled using autoregressive processes as discussed in modules 26 and 27.

So, if you remember in modules 26 and 27, we had discussed that autoregressive processes or autoregressive moving average processes show cyclical patterns. So, cyclical patterns are disturbing terms that are non-random, but disturbance as the term itself implies that they are supposed to be white noise processes, they are not explainable. So, we focus on the cyclical component or cyclical components that can be modelled.

(Refer Slide Time: 25:15)

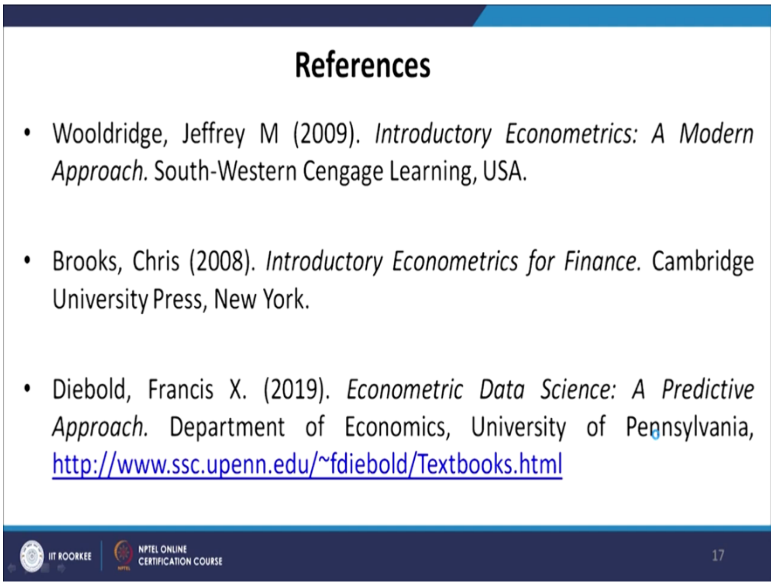


Therefore, a full model consisting of the trend seasonal and cyclical components could be written as (refer slide time: 25:22- 26:04). So, an ARP process will be having an either monotonically, geometrically monotonically decreasing per cycle, or maybe if there are some positive numbers, some negative numbers and we will be finding oscillating decay.

So, this implies that the regression is a complete regression model estimable using OLS, while the complete regression model looks like μ the constant path, this is the deterministic trend, this is the seasonal component, and this is the cyclical component plus this disturbance terms. So, this is how we can decompose a time series, we can estimate its individual components, either together or in consonance with each other.

And that is how univariate time series can be modelled, each and every one of its components can be modelled. This brings me to the end of the discussion on the univariate time series model.

(Refer Slide Time: 27:02)



These are the references that I have followed. Next time I will be taking up categorical variables or variables having multiple categories for both independent and dependent variables. Thank you.