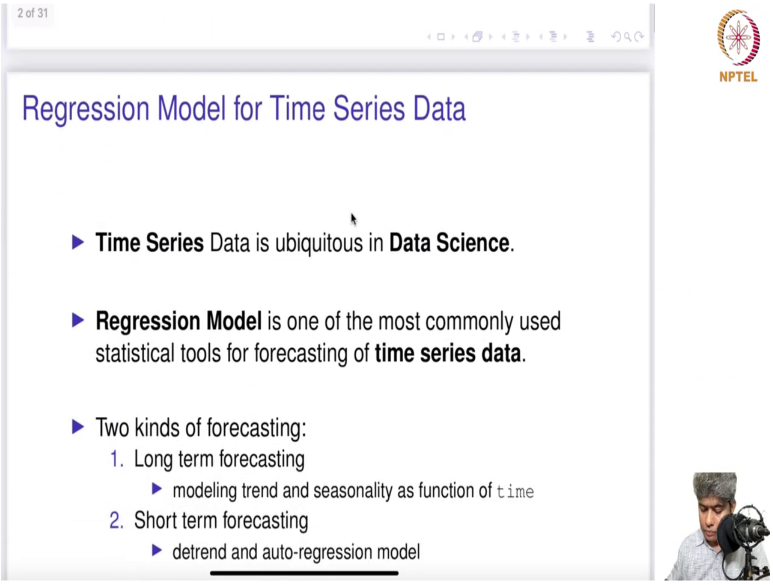
Predictive Analytics - Regression and Classification
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 26
Time Series Forecasting with Regression Model

Hello all, welcome to the part a of lecture eight, in this lecture we are going to discuss how we can use regression model simple regression model for time series data.

(Refer Slide Time: 00:34)



2 of 31

Regression Model for Time Series Data

- **Time Series** Data is ubiquitous in **Data Science**.
- **Regression Model** is one of the most commonly used statistical tools for forecasting of **time series data**.
- Two kinds of forecasting:
  1. Long term forecasting
     - modeling trend and seasonality as function of `time`
  2. Short term forecasting
     - detrend and auto-regression model

Time series data is very common in data science and whenever we see a data that comes with timestamp typically that is called time series data. So, far we have seen data likes empty curves data or diamonds data set; these data sets are not timestamp data.

So, these data set typically will be collected at a particular given point of time and these data sets are typically known as cross section data. In time series data what happens is each observations when it is collected will be the their time stamp like time minutes hours, day, which date of the year all those things all those information will be collected and as a one in a one column or several columns.

But overall, the time these kind of data called time set data or time series data. So, regression model is one of the most commonly used statistical tools for forecasting time series data ok. And we will see how can we use the regression model to forecast time series data; so, two kinds of forecasting depending on what is your objective. So, one is one objective is long term forecasting, you want to forecast in a long term.

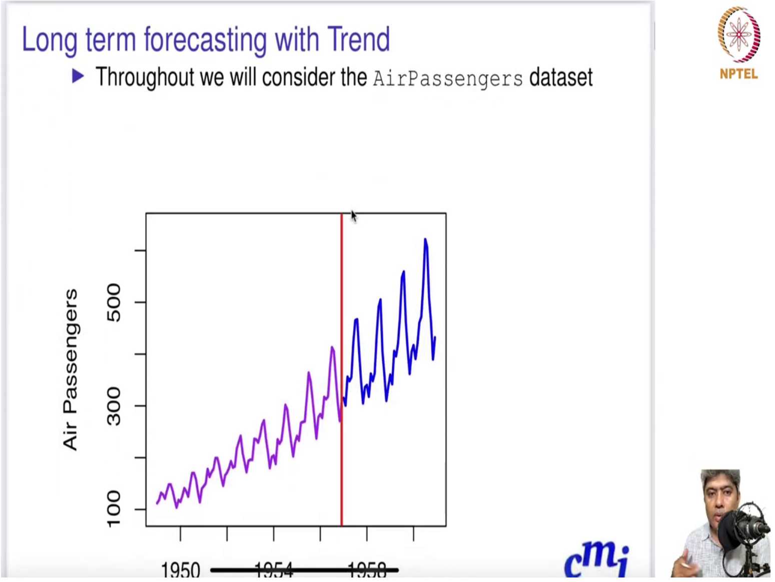
If you want to forecast long term, then modeling trained and seasonality as a function of time is very important and lot of effort we have to put in modeling trained and seasonality of the time series data. Another objective is short term forecasting; so, in short term forecasting we typically interested in forecasting for next time period, on next two time period that is it, maybe next three time periods that is it.

You will not even go up to four time period or five time points, but if you want to do a long time like you know quarterly forecasting or you know for a forecast for one year or two years ahead of time. Then that kind of forecasting will be a long term forecasting, but if your data is a monthly data and you just want to predict the next month or demand.

Or if it is a weekly data and just want to predict next week's demand, then it will be a typical called a short term forecasting. Just one time ahead or two time ahead forecasting will be considered as a you know short term forecasting, in that kind of forecasting we are not necessarily interested in the trained and seasonality that much. We are more interested in the local correlation and local feature of the data.

So, the in this case typically we detrained the data and auto regression models are very good in using this kind of you know if forecasting model. So, we will discuss that in both kind of modeling both kind of objective, long term forecasting and short term forecasting regression model can be used.

(Refer Slide Time: 04:17)



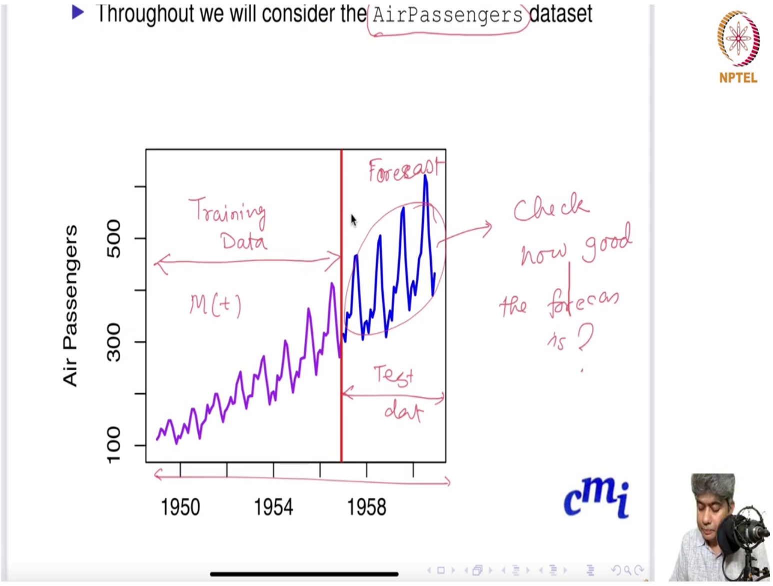
Long term forecasting with Trend

- Throughout we will consider the `AirPassengers` dataset

So, in order to understand long term forecasting with trend we will throughout, we will consider this air passenger data set, this is a small data set, but very good for understanding the concepts. As I discussed in my live session also that you know small data sets are good for understanding the small basic concepts and in the hands on we will use long and you know more complicated big data.

So, in this air passenger data set we have monthly number of people who traveled through in air in US between in I think 1949 to 1962 around this time period. And what we are going to do? We are going to take up to 1957 this period which is marked with purple as training data ok.

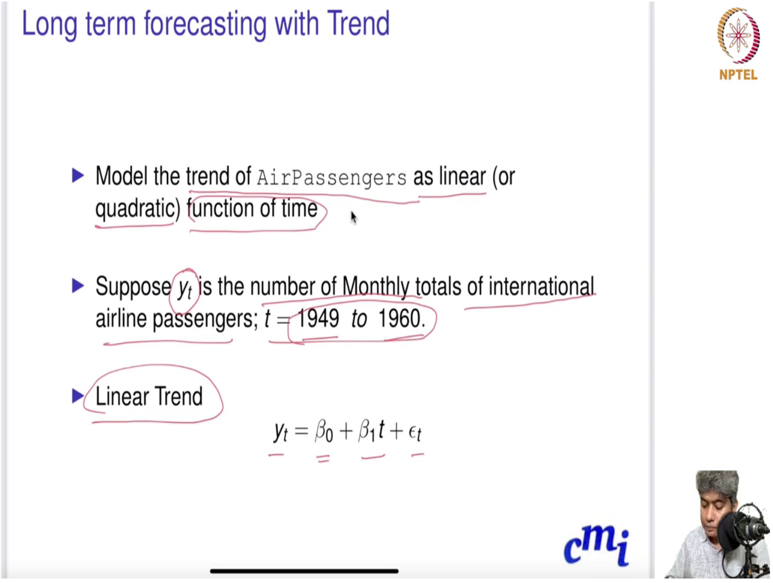
(Refer Slide Time: 05:09)



- Throughout we will consider the AirPassengers dataset

We are going to use it as a training data and this period we are going to use as test data ok. So, we will build our model based on this training data and then we will do the forecast here. We will do our forecast here and see how good our forecast will be and check how good the forecast is ok; so, this is our this will be our main objective.

(Refer Slide Time: 06:00)



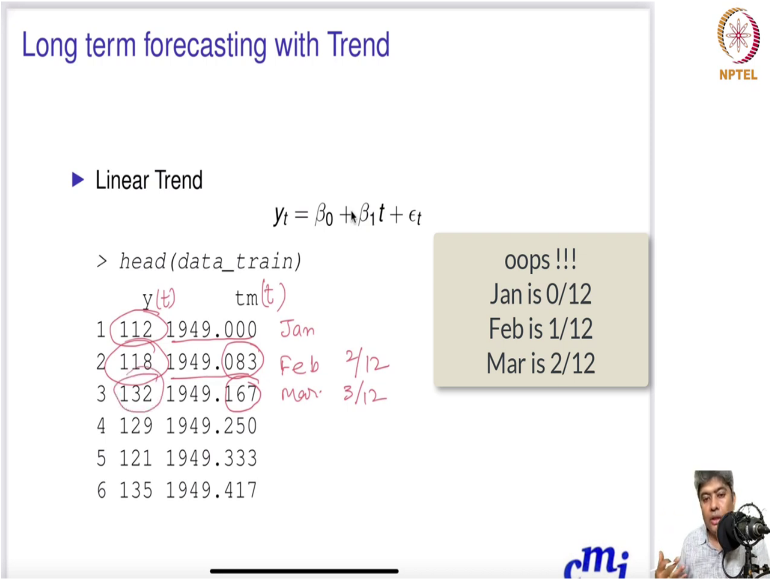
Long term forecasting with Trend

- Model the trend of `AirPassengers` as linear (or quadratic) function of time
- Suppose y_t is the number of Monthly totals of international airline passengers; $t = 1949 \;\; to \;\; 1960$.
- Linear Trend

$$y_t = \beta_0 + \beta_1 t + \epsilon_t$$

So, first thing is to model the trend of the air passengers data, first thing we will try maybe a linear or quadratic function as a function of time. So, suppose y t is the number of monthly totals of international airline passengers; so, that is the value and t runs from 149 to 1949 to 1960. So, it is a every during 19 from 1949 to 1960 every months how many international airline passengers were traveled that is being reported. So, if we think that there is a linear trend, what we will do? We will just fit a simple y t is equal to beta naught plus beta 1 plus epsilon t ok.

(Refer Slide Time: 07:00)



Long term forecasting with Trend

- Linear Trend

$$y_t = \beta_0 + \beta_1 t + \epsilon_t$$

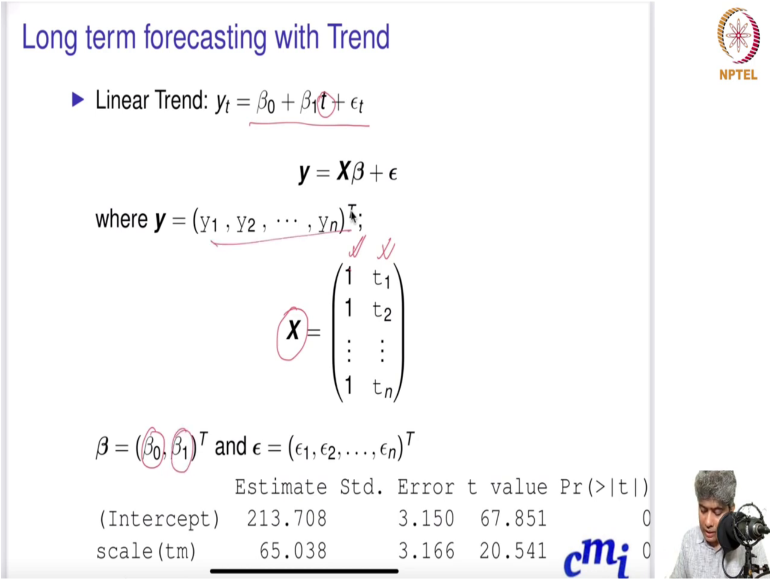
```
> head(data_train)
    y(t)      tm(t)
1  112  1949.000   Jan
2  118  1949.083   Feb   2/12
3  132  1949.167   Mar.  3/12
4  129  1949.250
5  121  1949.333
6  135  1949.417
```

oops !!!
Jan is 0/12
Feb is 1/12
Mar is 2/12

And if you see the data; so, you see 1949 this is Jan and this is Feb ok 1940 Feb of the 1949. So, what is happening is it is 2 out of 12 basically. So, that gives you 0.83 and this is March, a March is kind of 3 out of 12 that has given you 1949.167; so, that is how the months are being converted into a numeric value ok.

So, this is my t or tm simple t or tm and this is the y of t essentially, these are the value these are these many people or these many in in 100000's 112000 and 18000, 132000 in that month people have traveled through air.

(Refer Slide Time: 07:58)



Long term forecasting with Trend

- Linear Trend: $y_t = \beta_0 + \beta_1 t + \epsilon_t$

$$\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon$$

where $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)^T$;

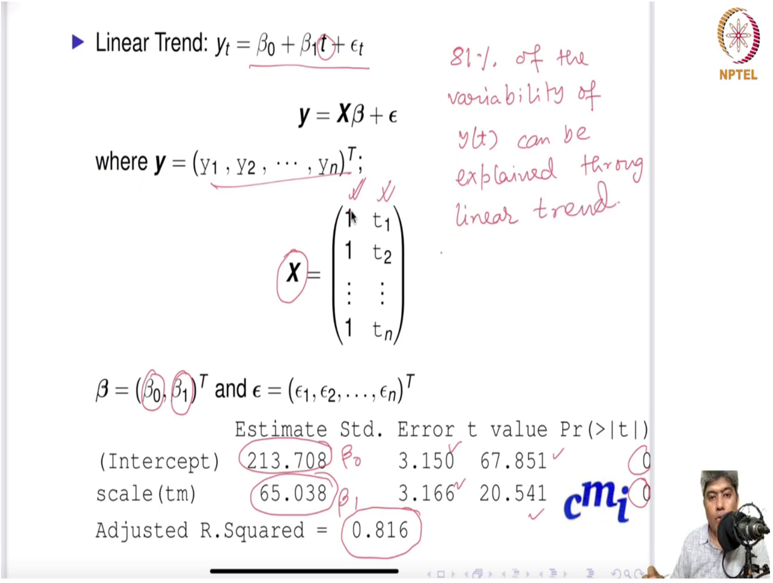
$$\boldsymbol{X} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}$$

$\beta = (\beta_0, \beta_1)^T$ and $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  213.708      3.150  67.851        0
scale(tm)     65.038      3.166  20.541        0
```

So, we can just simply put them into a simple linear regression framework and our design matrix. So, it will be 1 1 1 on the first column it will be intercept, second column will be t and beta will be beta naught and beta 1 this is y 1 to yn.

(Refer Slide Time: 08:26)



- Linear Trend: $y_t = \beta_0 + \beta_1 t + \epsilon_t$

$$\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon$$

where $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)^T$;

$$\boldsymbol{X} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}$$

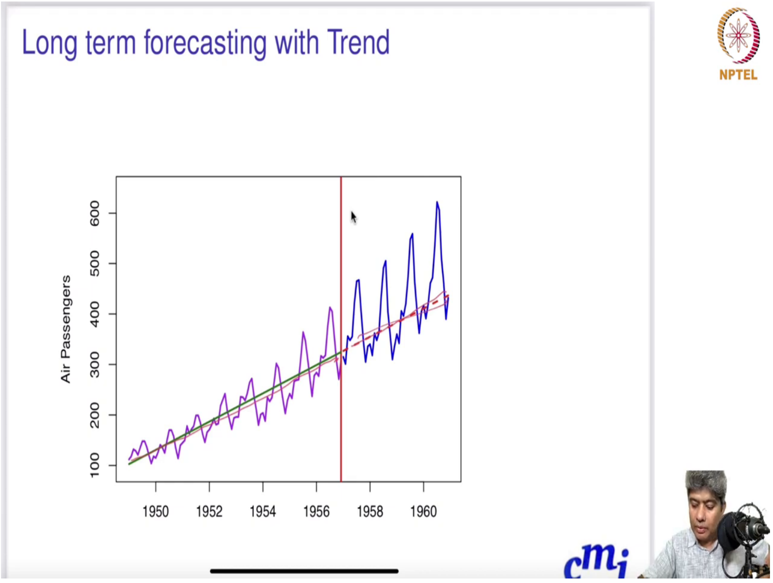
$\beta = (\beta_0, \beta_1)^T$ and $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   213.708  β0   3.150   67.851        0
scale(tm)      65.038  β1   3.166   20.541        0
Adjusted R.Squared =  0.816
```

81% of the variability of y(t) can be explained throug linear trend

And now if you fit this model, this will be the final you know estimates of the model this is beta naught and this is beta 1. These are the standard data, these are the t value and the p value is so small that it is pretty much rounded up with 0; so, adjusted r square tons out to be 816. So, almost 81 percent of the entire variability of the data can be explained through the trend. So, 81 of the variability of the data can be explained through the trend; so, 80 81 percent of the variability ability of y t can be explained through linear trend ok.

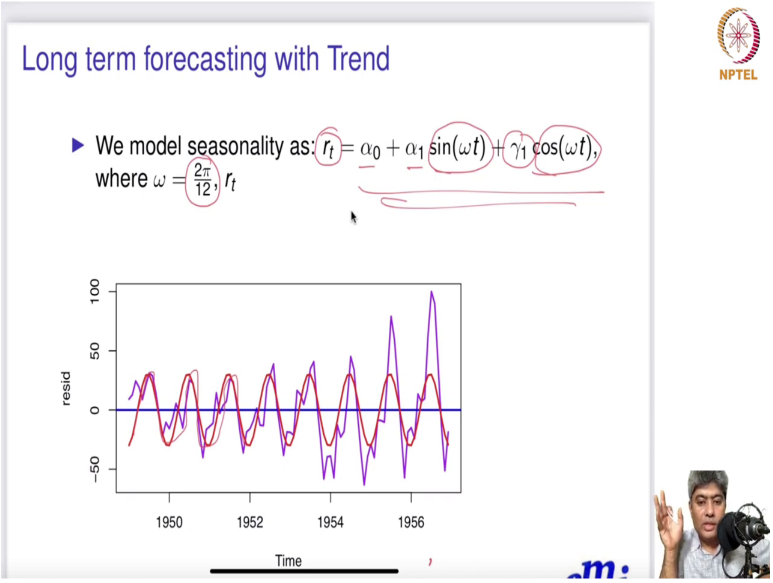
(Refer Slide Time: 09:36)



Long term forecasting with Trend

And if you fit this simple linear trend this is the simple linear trend and this is the forecast. So, clearly it can pick up the trend model the trend correctly, but if you plot the residuals clearly residuals pick up the seasonality ok.

(Refer Slide Time: 09:51)



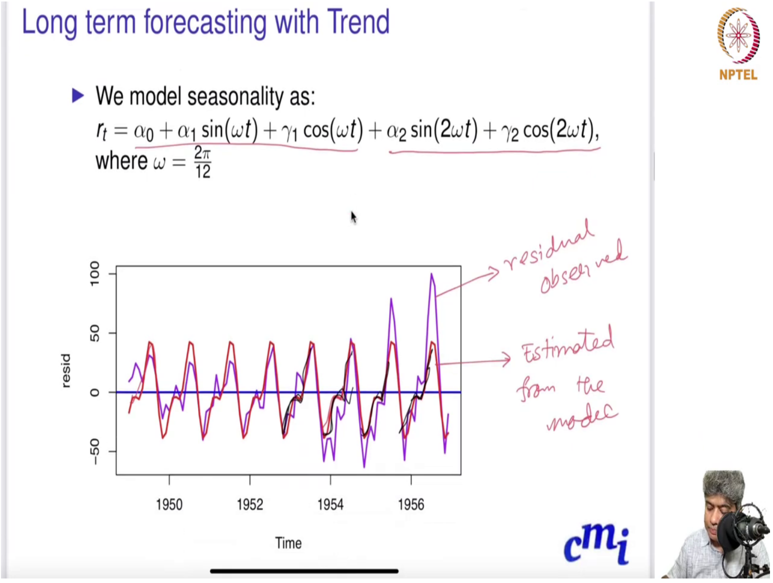
Long term forecasting with Trend

- We model seasonality as: $r_t = \alpha_0 + \alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t)$, where $\omega = \frac{2\pi}{12}$, r_t

Now, we have to model the seasonality we have to model the seasonality; so, how we model seasonality? So, we can model seasonality r t residual as a function of alpha naught plus alpha 1 sin omega t plus gamma 1 cos omega t. Now, what is omega? Omega we are going to take 2 pi by 12, since we have a year; so, we can take a 2 pi by 12.

And if we fit this model simple model with Fourier transform, you can see that you know simple one Fourier transform can be modeled model the residual 1 and 2 greater extent. But you can see that these models are not these are the places where it is not very well captured.

(Refer Slide Time: 11:04)



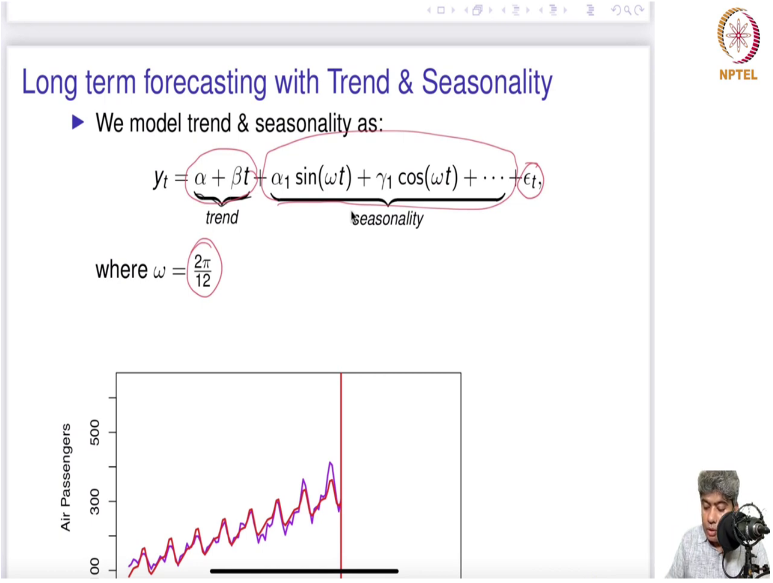
Long term forecasting with Trend

- We model seasonality as:
  $r_t = \alpha_0 + \underline{\alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t)} + \underline{\alpha_2 \sin(2\omega t) + \gamma_2 \cos(2\omega t)}$,
  where $\omega = \frac{2\pi}{12}$

Now, what about after we stopped after first Fourier transform what about we are taking the second Fourier transform. So, we can just after that we can just alpha 2 sin 2 omega 2 plus 2 omega 2 and you can clearly see this red line, purple line is the real the residual through residual or residual observed, and these red line is the estimated from the model.

So, clearly when you put the second Fourier it tries to you know capture this curve you know. It tries to capture this particular chain, you know this seasonality, this summer it is I think this is the summer kind of before summer kind of pick a curve that is getting picked up.

(Refer Slide Time: 12:20)



Long term forecasting with Trend & Seasonality

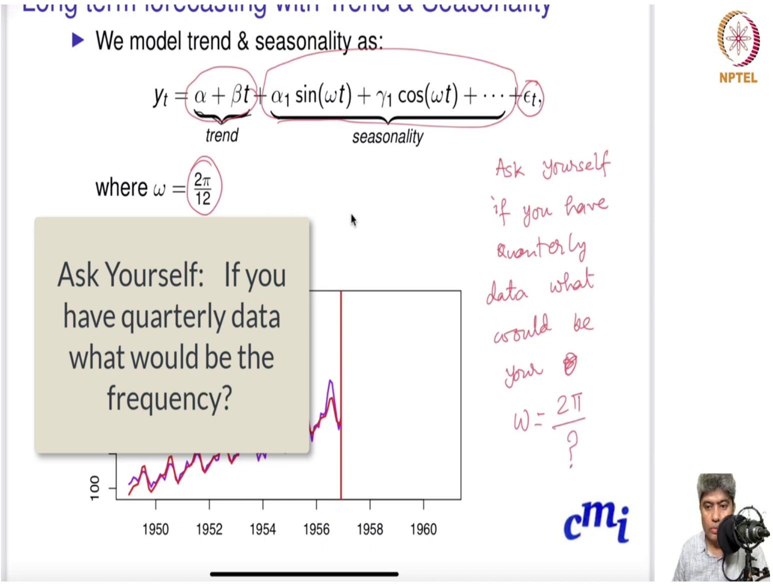
- We model trend & seasonality as:

$$y_t = \underbrace{\alpha + \beta t}_{trend} + \underbrace{\alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t) + \cdots}_{seasonality} + \epsilon_t,$$

where $\omega = \frac{2\pi}{12}$

So, now we if we join the trend and seasonality, we define our model alpha plus beta t trend alpha plus beta t; yes, trend this is the trend part and this is the seasonality part. You can the seasonality you can put as many Fourier transform as you want and plus epsilon t of course. And omega is 2 pi by 12, because we want it is a monthly data that we have; so, therefore, the frequency is 12.

(Refer Slide Time: 12:57)



Long term forecasting with Trend & Seasonality

- We model trend & seasonality as:

$$y_t = \underbrace{\alpha + \beta t}_{trend} + \underbrace{\alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t) + \cdots}_{seasonality} + \epsilon_t,$$

where $\omega = \frac{2\pi}{12}$

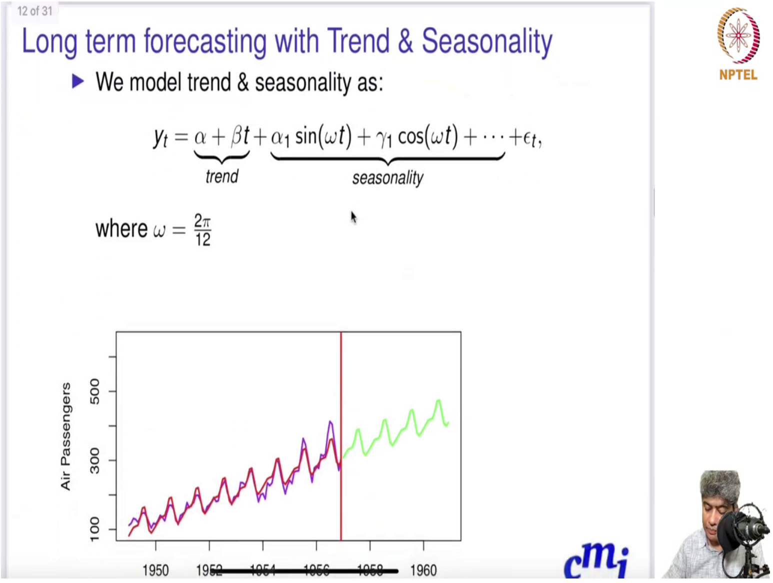
Ask Yourself: If you have quarterly data what would be the frequency?

Ask yourself if you have quarterly data what would be your ω

$\omega = \frac{2\pi}{?}$

If we had quarterly data what would be the frequency just figure out ask yourself, take a break maybe you know pause your video for five minutes ask yourself. Ask yourself, if you have quarterly data quarterly data, data what would be your omega of frequency; so, 2 pi by what which value you should use alright.

(Refer Slide Time: 13:48)



12 of 31

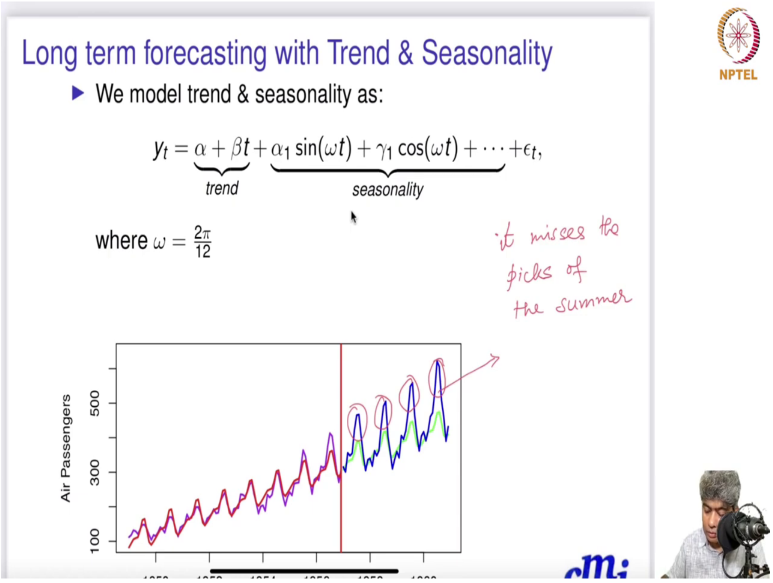
Long term forecasting with Trend & Seasonality

- We model trend & seasonality as:

$$y_t = \underbrace{\alpha + \beta t}_{trend} + \underbrace{\alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t) + \cdots}_{seasonality} + \epsilon_t,$$

where $\omega = \frac{2\pi}{12}$

(Refer Slide Time: 14:08)



Long term forecasting with Trend & Seasonality

- We model trend & seasonality as:

$$y_t = \underbrace{\alpha + \beta t}_{trend} + \underbrace{\alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t) + \cdots}_{seasonality} + \epsilon_t,$$

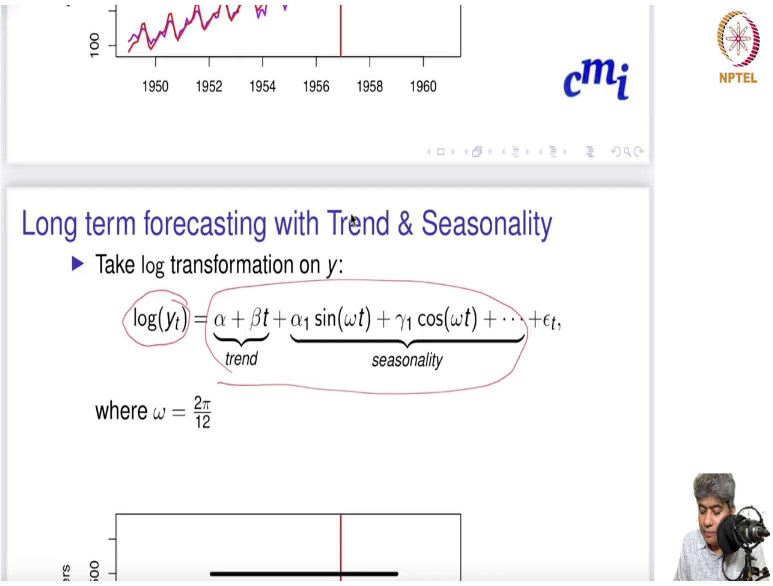
where $\omega = \frac{2\pi}{12}$

it misses the picks of the summer

Next so, if you use this if you fit this model and if you forecast this this green part, this green thing then forecast it kind of gives. And then on the top of that if you just blue one is the actual data, it picks up the seasonality, but it clearly misses all the picks in the summer, it clearly misses all the picks of the summer it misses the picks of the summer.

Now, in this situation we I thought we thought like you know a transformation might help, because if you look it carefully it looks like these picks are increasing at an exponential rate.

(Refer Slide Time: 14:58)



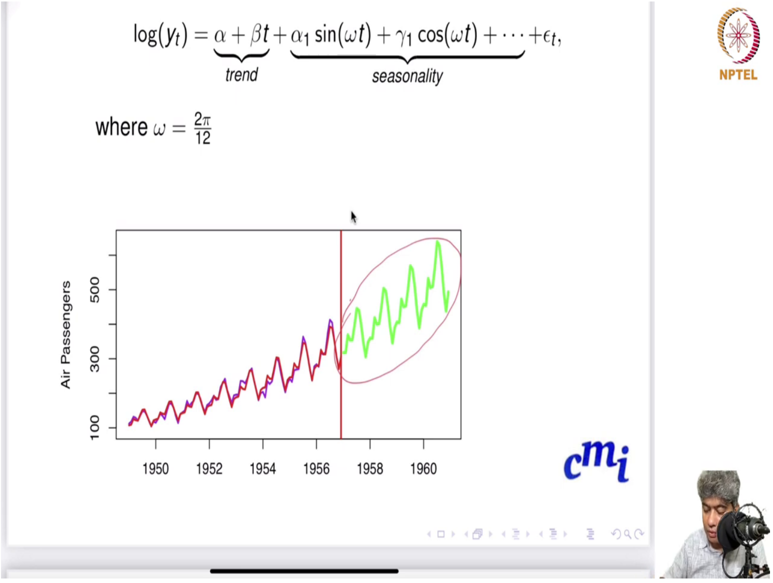
Long term forecasting with Trend & Seasonality

- Take log transformation on y:

$$\log(y_t) = \underbrace{\alpha + \beta t}_{trend} + \underbrace{\alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t) + \cdots}_{seasonality} + \epsilon_t,$$

where $\omega = \frac{2\pi}{12}$

(Refer Slide Time: 15:13)

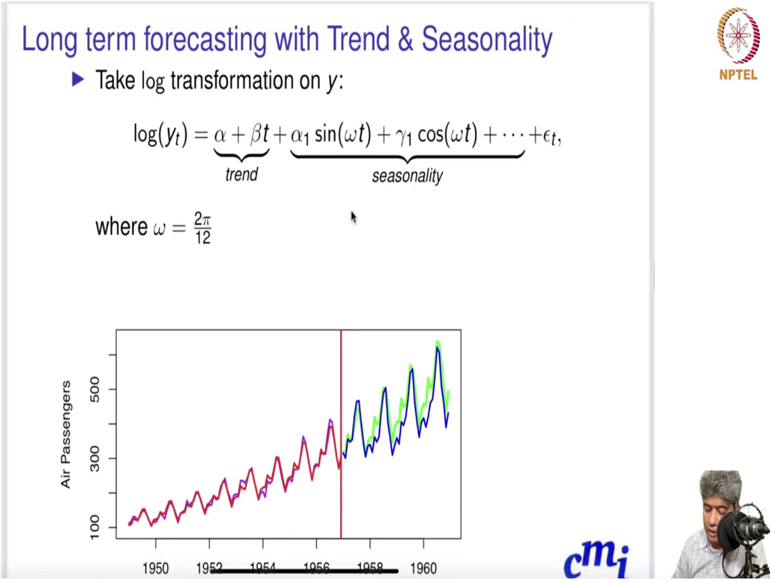


$$\log(y_t) = \underbrace{\alpha + \beta t}_{trend} + \underbrace{\alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t) + \cdots}_{seasonality} + \epsilon_t,$$

where $\omega = \frac{2\pi}{12}$

So, no harm try to model it as exponential transformation after a log transformation. So, I keep the right side of the model exactly same, but on the left side instead of modeling y t, I used I took log of y t. And when I did that looks like during the fitting part is doing pretty decent and in the forecast also looks like they have it has done quite decent work in the out of the sample forecast also it has done quite full work And then when we put the real values it is almost matching exactly.

(Refer Slide Time: 15:30)



Long term forecasting with Trend & Seasonality

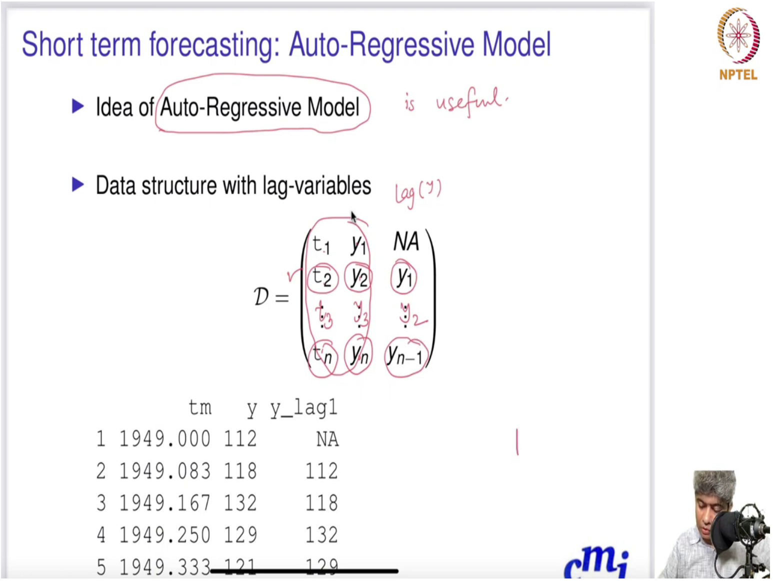
- Take log transformation on y:

$$\log(y_t) = \underbrace{\alpha + \beta t}_{trend} + \underbrace{\alpha_1 \sin(\omega t) + \gamma_1 \cos(\omega t) + \cdots}_{seasonality} + \epsilon_t,$$

where $\omega = \frac{2\pi}{12}$

So, when you and here it is we are considering these are all long term forecasting for 1, 2, 3, 4 almost 4 years we are forecasting the trend and the seasonality and it is doing pretty decent job.

(Refer Slide Time: 16:00)



Short term forecasting: Auto-Regressive Model

- Idea of Auto-Regressive Model is useful.
- Data structure with lag-variables lag(y)

$$\mathcal{D} = \begin{pmatrix} t_1 & y_1 & NA \\ t_2 & y_2 & y_1 \\ t_3 & y_3 & y_2 \\ t_n & y_n & y_{n-1} \end{pmatrix}$$

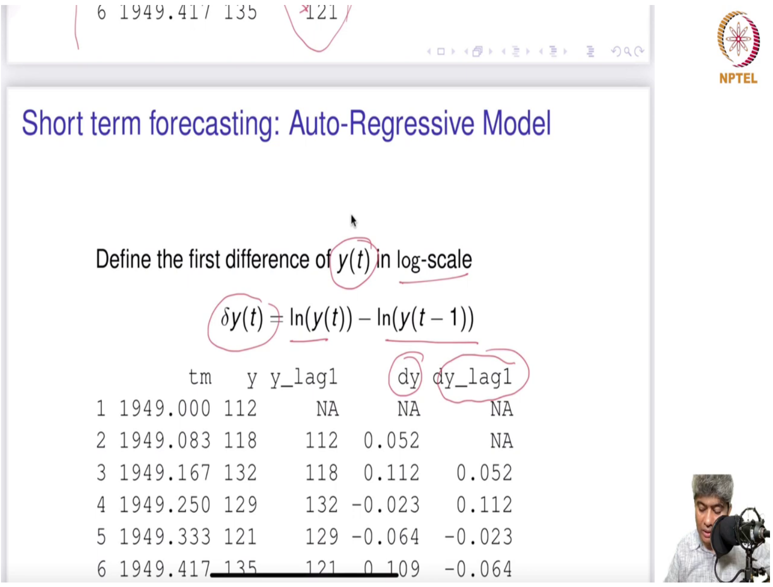
```
        tm   y y_lag1
1 1949.000 112     NA
2 1949.083 118    112
3 1949.167 132    118
4 1949.250 129    132
5 1949.333 121    129
```

Now, next we will move on to modeling short term forecasting, here our objective is focused only 1 month ahead or maybe 2 month ahead not more than that. In this kind of situation idea of auto regressive model is very useful, auto regressive model is useful it is very useful; so, how the data structure works.

So, what you have to do? You have to just create a bunch of lag variables; so, we will start with one. So, this is my data actually this is my data t 2, t 2, t n, y 1, y 2, y n and then I just take create a lag of y, what is lag of y? So, in the time point 2 time point 2, y 2 is being observed.

And what is the one variable lag variable? So, that is y 1; so, that is y 1. Similarly, if you have a t 3 y 3, what will be the lag variable? y 2 sim and the n t n y n the lag variable is y n minus 1.

(Refer Slide Time: 17:23)



```
6 1949.417 135   121
```

Short term forecasting: Auto-Regressive Model

Define the first difference of $y(t)$ in log-scale

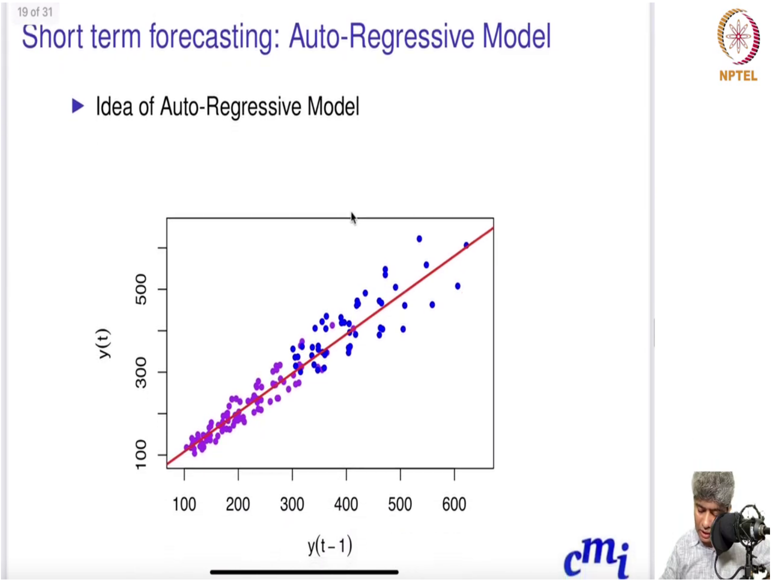
$$\delta y(t) = \ln(y(t)) - \ln(y(t-1))$$

```
        tm   y y_lag1     dy dy_lag1
1 1949.000 112     NA     NA      NA
2 1949.083 118    112  0.052      NA
3 1949.167 132    118  0.112   0.052
4 1949.250 129    132 -0.023   0.112
5 1949.333 121    129 -0.064  -0.023
6 1949.417 135    121  0.109  -0.064
```

So, you can you create a lag variable in the data set also we created a lag variable. So, just by placing in this way we created a lag variable and our plan is and how what you can do is essentially take the differences of the lag variable. So, define the first difference of the y t in log scale as log y t minus log of y t minus 1, because log y t is our we found that you know in the long run forecasting was very good.

So, we just take the differs differences of the you know; so, d y is essentially the d y is the essentially the this guy, we take the log y and different take the consecutive differences. And then you can take the lag of even the this the first difference order variable as well; so, this is how you create the lagged variable.

(Refer Slide Time: 18:34)

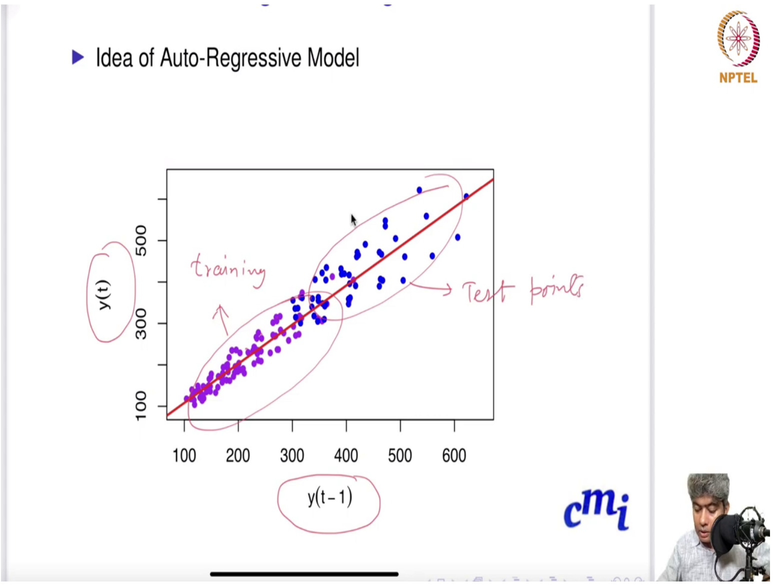


19 of 31

Short term forecasting: Auto-Regressive Model

- Idea of Auto-Regressive Model

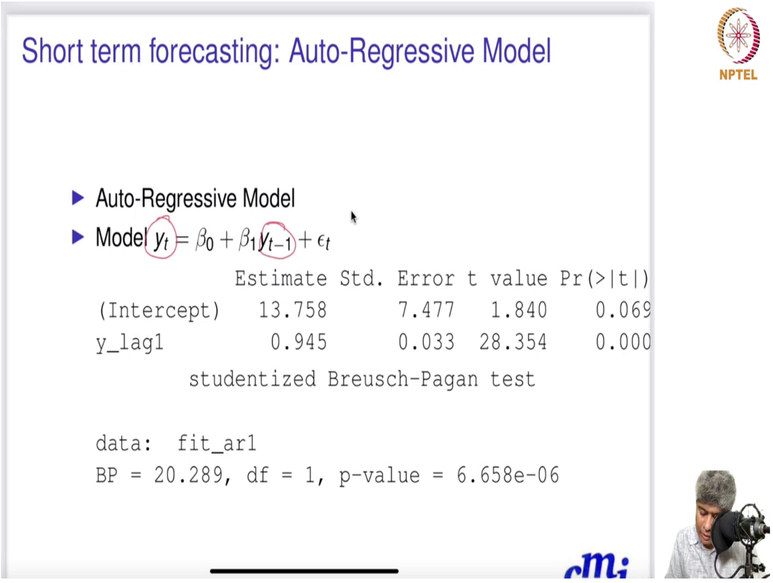
(Refer Slide Time: 19:03)



- Idea of Auto-Regressive Model

Now, you see if you just plot on the y axis y t minus 1 and the y t, you can clearly see that there is a increasing trend. But what is happening and this particles are all training data points and blues are all test data particles, these purples are training points training points and these blues are tests points ok.

(Refer Slide Time: 19:14)



Short term forecasting: Auto-Regressive Model

- Auto-Regressive Model
- Model $y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$

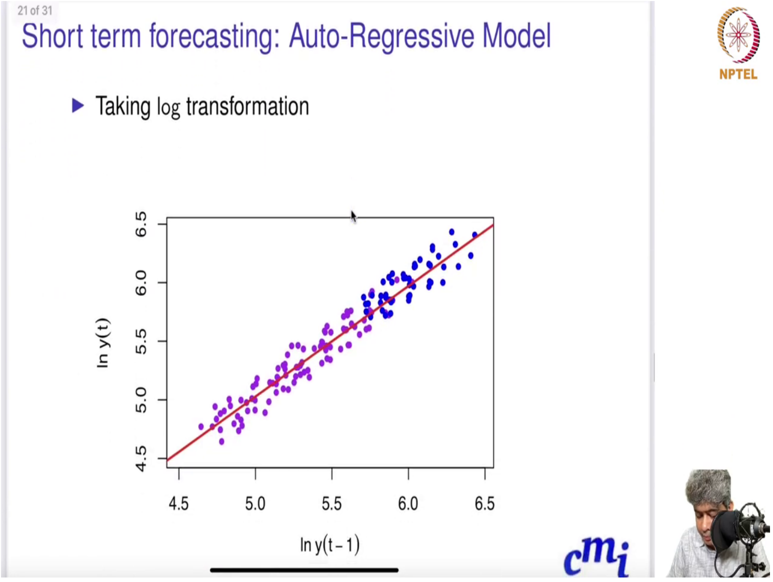
```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   13.758      7.477   1.840    0.069
y_lag1         0.945      0.033  28.354    0.000
        studentized Breusch-Pagan test

data:  fit_ar1
BP = 20.289, df = 1, p-value = 6.658e-06
```

And if you fit the you know auto regressive model of y t as a function of y t minus 1 and you can do after fitting the model it did a breusch pagan test on the residual and the p value is very small. I am not surprised p value is very small; that means, there is a being bunch of heteroscedasticity, because you see at the beginning there is a kind of tight in the training data was kind of run tight, but the test data was kind of data out.

So, the there is; so, the residuals are not homogeneous residuals are heterogeneous. So, I am not surprised that breusch pagan test will reject the header homoscedasticity and it will go for heteroscedasticity.

(Refer Slide Time: 20:05)



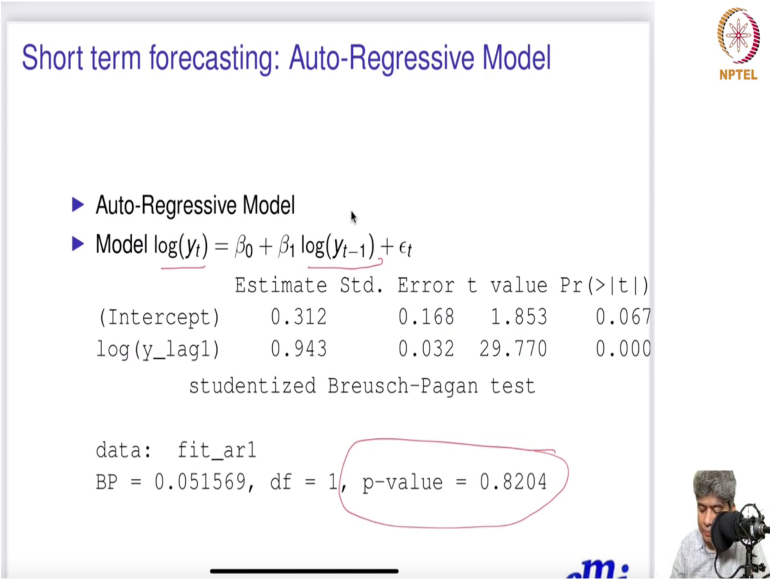
21 of 31

Short term forecasting: Auto-Regressive Model

- Taking log transformation

So, I have taken the log y t minus 1 and log y t and splot them. And now it looks like they are somewhat homogeneous behavior, there is a somewhat homogeneous behavior which is probably correct.

(Refer Slide Time: 20:22)



Short term forecasting: Auto-Regressive Model

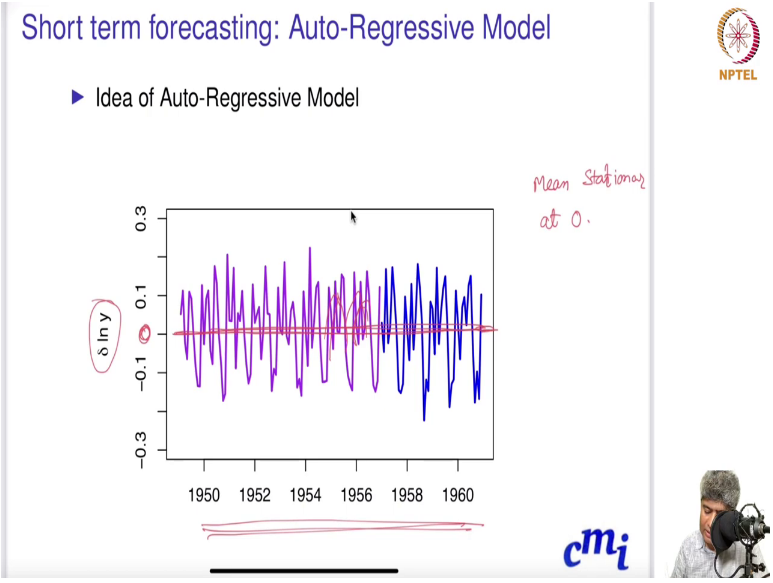
- Auto-Regressive Model
- Model $\log(y_t) = \beta_0 + \beta_1 \log(y_{t-1}) + \epsilon_t$

```
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.312      0.168    1.853    0.067
log(y_lag1)      0.943      0.032   29.770    0.000
           studentized Breusch-Pagan test

data:  fit_ar1
BP = 0.051569, df = 1, p-value = 0.8204
```

Now, if you fit this log y t beta naught plus beta 1 log y t minus 1, you see that breusch pagan test p value is very high so; that means, it cannot reject the null hypothesis that it is homoscedastic residuals or homoscedastic. So, the residuals are homoscedastic that assumption is correct and this is a indeed a good starting point for lag one auto regressive model.

(Refer Slide Time: 21:00)



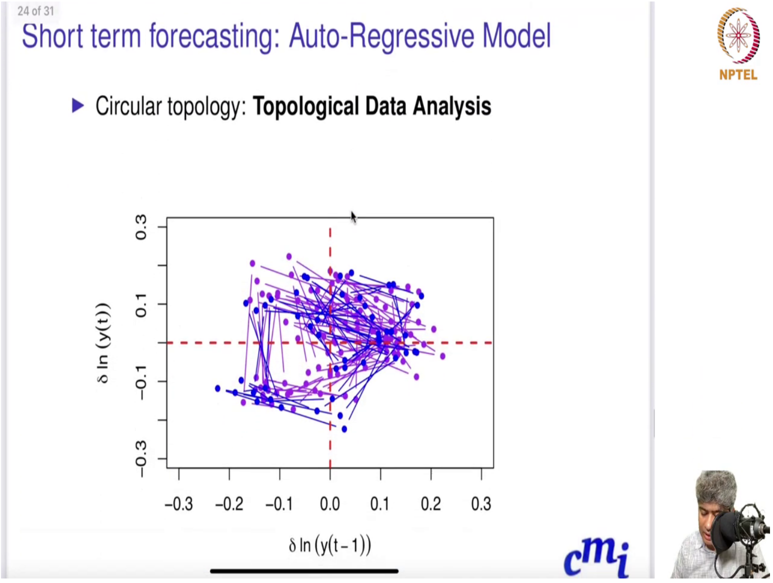
Short term forecasting: Auto-Regressive Model

- Idea of Auto-Regressive Model

Mean Stationary at 0.

So, what is the idea of auto regressive model? If you basically now fit this fit this you know plot this delta of log y over the time period horizon this of 1950 to 1960. Then what you will see that this differences is essentially mean stationary, it essentially means stationary it is a mean stationary where it means stationary at 0 ok.

(Refer Slide Time: 21:55)



24 of 31

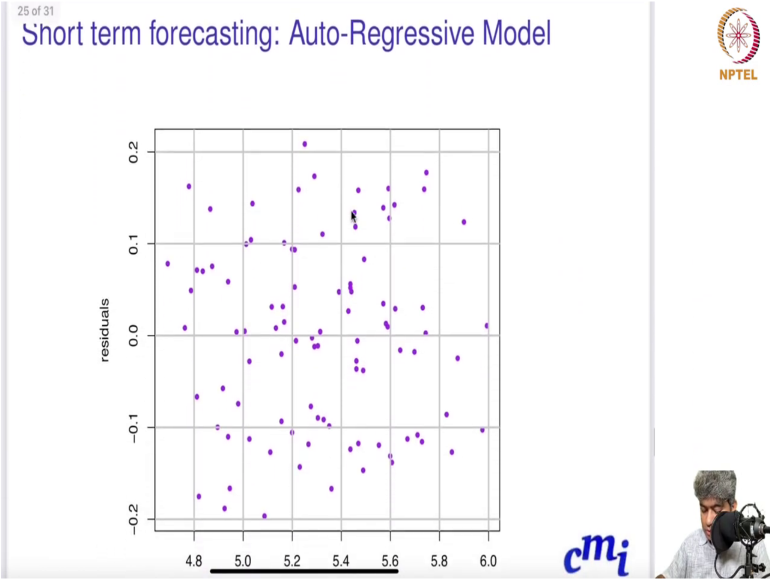
Short term forecasting: Auto-Regressive Model

- Circular topology: **Topological Data Analysis**

So, this is the first thing, but still, we can see there is bunch of seasonality we can see. We will talk about it how can you do that and then here what I have done I have plotted the delta log y t minus 1 and delta log y t. And what we are finding is there is a circular hole in the data ok, and this is x effectively shows that how the seasonality is affecting in the there is still good decent seasonality.

And this is typically called in the circular topology and this falls into the category of topological data analysis. So, I am not going to discuss about this, but topological data analysis is a new top area and it is a you know very very exciting area that what is the topology of the data understand, is there a hole in the data and what we are seeing that there is indeed a hole in the data.

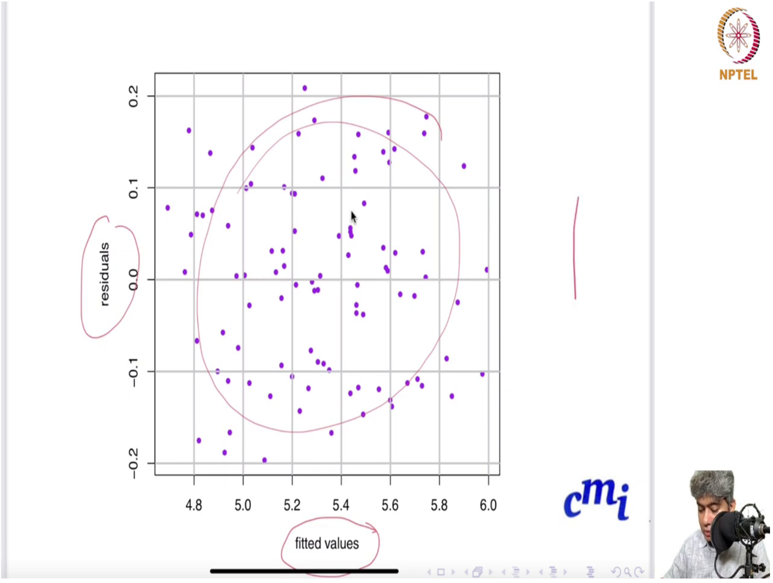
(Refer Slide Time: 22:52)



25 of 31

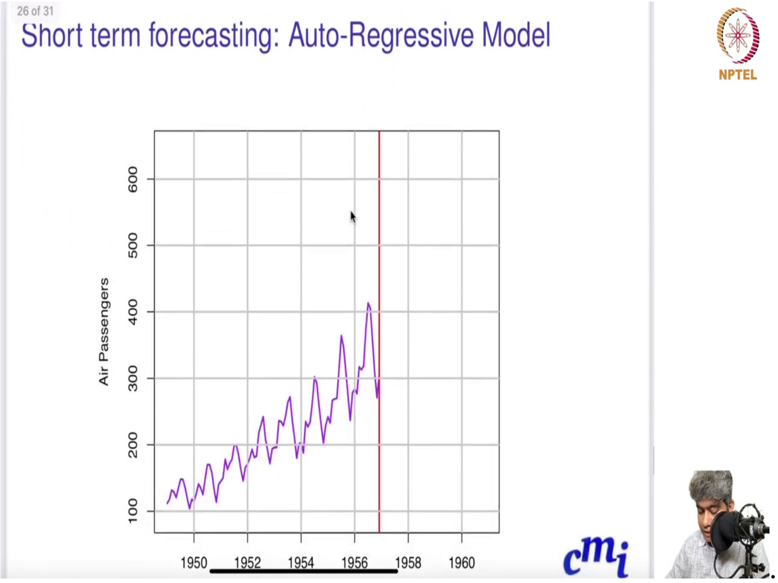
Short term forecasting: Auto-Regressive Model

(Refer Slide Time: 22:54)



So, here I have plot the fitted values against the residuals, what we are seeing that there is not much is going on which is a good news for us.

(Refer Slide Time: 23:04)

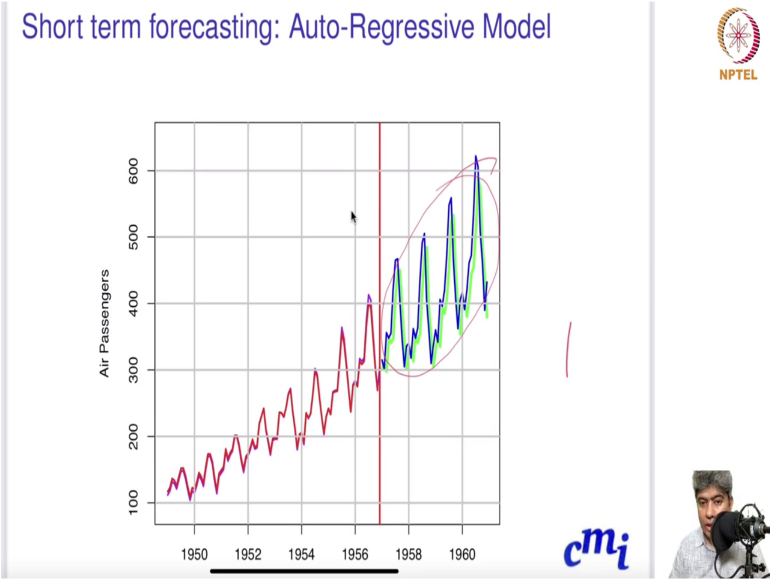


26 of 31

Short term forecasting: Auto-Regressive Model

And then when we took the data plotted it, and what we found that autoregressive model is almost one order autoregressive model was on almost mimicking the data.

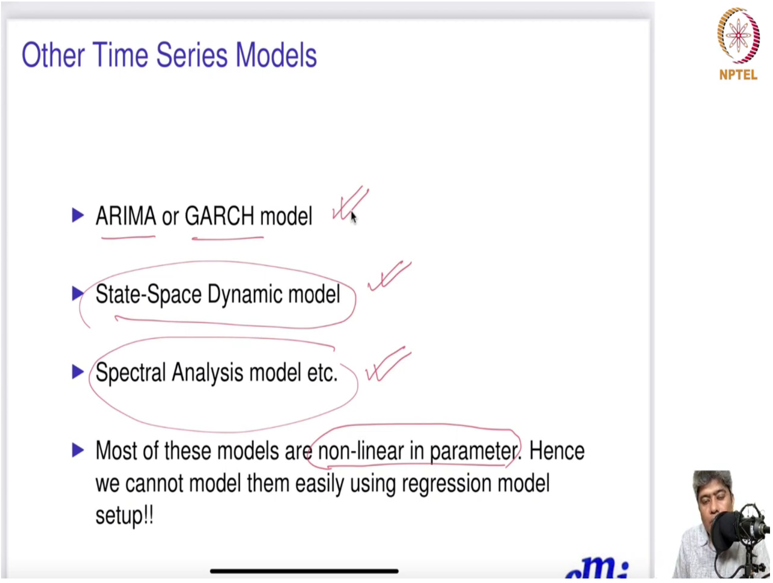
(Refer Slide Time: 23:20)



Short term forecasting: Auto-Regressive Model

Because it is just forecasting only in the next step; so, and in the out of the sample also it is almost mimicking the data; so, it is pretty good in terms of forecasting the model.

(Refer Slide Time: 23:33)



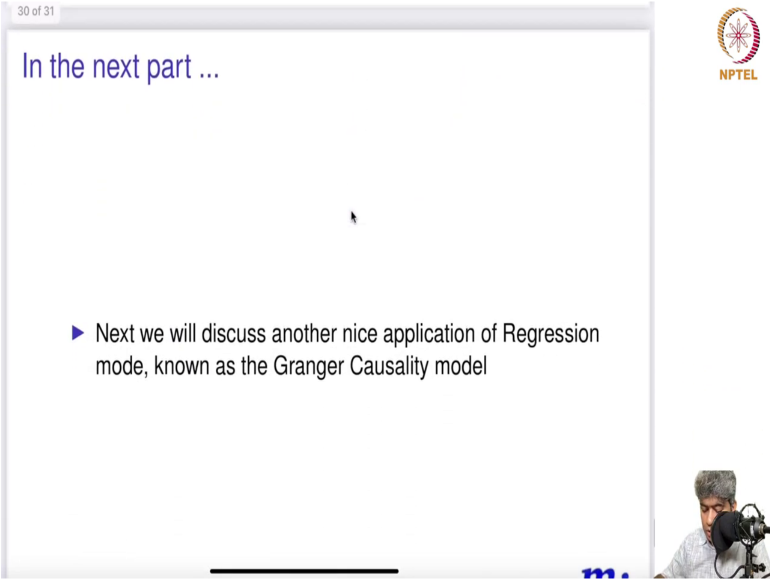
Other Time Series Models

- ARIMA or GARCH model
- State-Space Dynamic model
- Spectral Analysis model etc.
- Most of these models are non-linear in parameter. Hence we cannot model them easily using regression model setup!!

So, other time series models such as ARIMA model, GARCH model, State-Space Dynamic models, Spectral Analysis model most of these models are non-linear in parameter. So, simple trend and seasonality model for long term forecasting and auto regressive model for seasonality, for short term forecasting is regular regression model with the which is linear in parameter.

But these models these are very popular time series model these are not linear in term parameter they are non-linear in parameter. So, these models are right now beyond the you know periphery of this course, you we have to when we I will offer a full course on time series modeling. Of course, I will discuss all this model in that course, hence we cannot model them easily using regression model setup ok.

(Refer Slide Time: 24:35)



In the next part ...

- Next we will discuss another nice application of Regression mode, known as the Granger Causality model

So, I will stop here next we will do a hands on and then then we will discuss another nice application of regression model known as granger causality model; so, for now I will stop here.

Thank you see you in the next video.