Data Analysis and Decision Making - 3
Professor Raghu Nandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology Kanpur, India

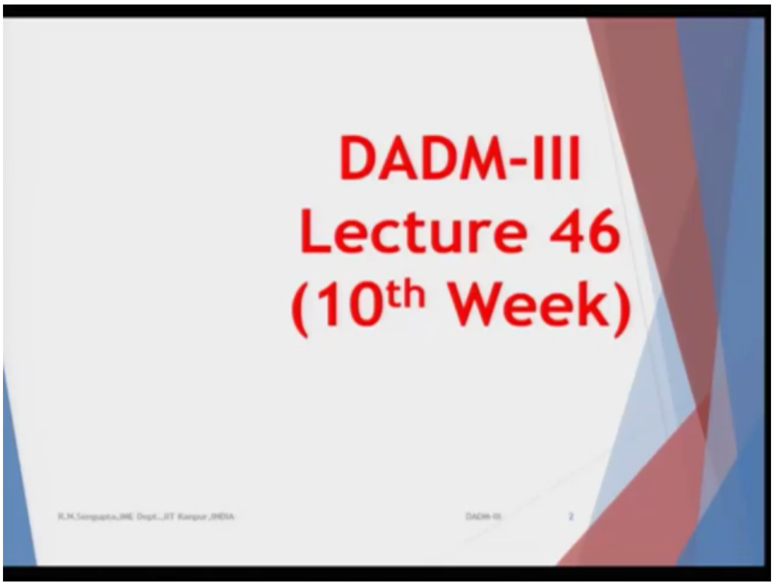
Lecture 46

Welcome back, my dear friends, a very good morning, good afternoon, good evening to all of you. And as you know this is the DADM-3 course, which is Data Analysis and Decision Making course under the NPTEL MOOK series. And this total course duration, contact hours is 30 hours which when spread over number of lectures is 60 because each course or each lecture is for half an hour and the total number of weeks it is spread over is 12.

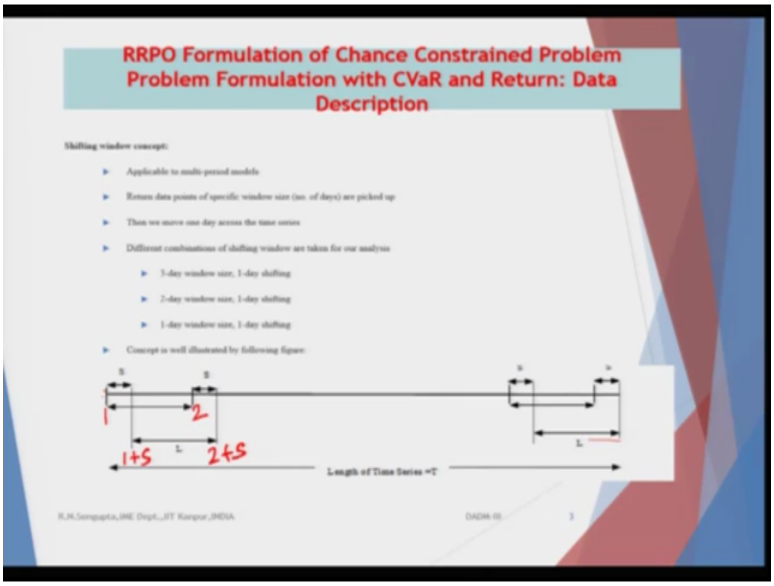
And we have already completed 45 lectures that means we have completed 9 weeks, we will be starting the 10^{th} week. And after each week, where each week we have 5 lectures, you have one assignment you have already completed in 9 assignments. By the time you are going to hear me about this 10^{th} week and after the 12 assignments you will take the final examination. And my good name is Raghu Nandan Sengupta from the IME department at IIT Kanpur.

So if you remember we were discussing about, the optimization quadratic programming I did mention that we will consider the branch and bound method with an example and then we will go into the other realms of reliability and robust programming. So I just giving you a scenario or an analysis how the graphs and the outputs will look like we then will come to the detail analysis in the reliability and robust concept.

(Refer Slide Time: 1:44)



(Refer Slide Time: 1:47)



So as you can see from the slide this is the DADM-3 lecture number 46, which we are starting in the 10th week. Now whenever we are considering the optimization from mainly from the perspective for this problem for portfolio analysis. Portfolio means you have a basket where different stocks are there, you want to optimize them and when you want to optimize them that there would be basically the returns which will be found from the expected value. There will be a risk which will be found from the variance.

And you have to basically optimize using the first moment or the second moment considering the fact that you are trying to utilize the concept of utility analysis where you will try to basically maximize the expected utility, which is the mean and try to basically minimize the risk of the utility which is the variance. And when you basically do that, you are trying to basically for consider the portfolio as such, so you have to basically consider a conglomeration of such assets.

So this model which you are considered it is not dynamic in nature but we will consider that it will be applicable for the whole time period, again we will consider the same thing. We will consider the total length of the time series is T, capital T and we will break it into parts, parts means either a week or 2 weeks and take the average of these 2 weeks, so each week will have you will have 5 working days.

Based on those 5 working days for the 2 weeks you will have 10 number of working days, 10 number of working days when converted you will try to find out the average value and that average value will, will also be consider a random variable because you are trying to take the average of the samples and then proceed accordingly such that you are able to find and overlapping windows which will give you much better analysis and results considering that you are concentrating both on the in-sample and the out-sample.

So shifting window concept will be is generally used in econometric since time series in finance in in all the domains in this area but it is applicable to multi period models, so you are basically trying to consider the multi at different points of time what is the output and consider their analysis the models, actual results based on the data we will consider the data as in-sample, where you try to basically analyze the problem for the first time and then try to utilize it in the out-sample.

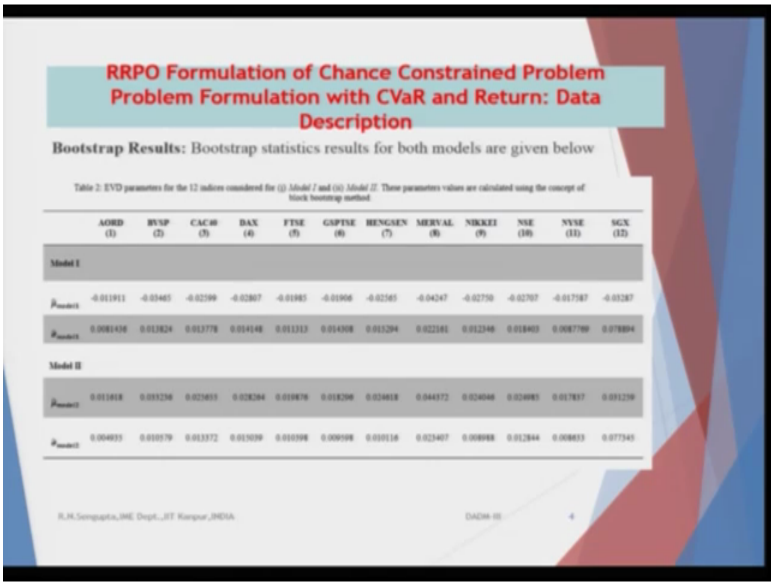
You will basically find the return data of the specific in window, we will be picked up then we will move 1 day at a time or 3 days at a time or 4 days a time depending on how many number of overlaps you want and different combinations can be considered. You can consider 3 day window, which I just mentioned few minutes back. If it is 2 weeks, 2 weeks would basically mean 10 days, so it is a 3 day window with 1 day shifting, so 1 shift would be too little it can be say for example 3, 5 day shifting also considering that you are considering 2 weeks as a window.

So that means 2 weeks shift (1 day) 1 week, 2 weeks shift 1 week and then you proceed accordingly. So few of the combinations which are just stated, which can be tried would be 3 days window size, which was 1 day shifting, 2 days window size is 1 day shifting, 1 day window sides with 1 day shifting and so on and so forth. So concept is basically illustrated in the following figure, where you have the time frame, so this is the window the total length is starting from 0 till capital T and these are the lengths of the windows L.

They all have the lengths of the windows and it shifts, so shifts are given by S. So the first window is basically considered from 1 to 2, the second window if it shifts by S days, it will be 1 plus S to 2 plus S and then correspondingly it will shift. So the third window if it shifts, it will shift by 1 plus S plus S. So hence, the other end of the window will also shift by 2 plus S plus S and it will continue accordingly. And you will find out the overall performance of the model for these shifting windows.

So these shifting windows are they are happening for each and every stock, so will try to find out the average, find out the variances of this these window shifts then find out the correlation coefficient and then they try to find out the weights based on which you can model the overall portfolio.

(Refer Slide Time: 6:24)

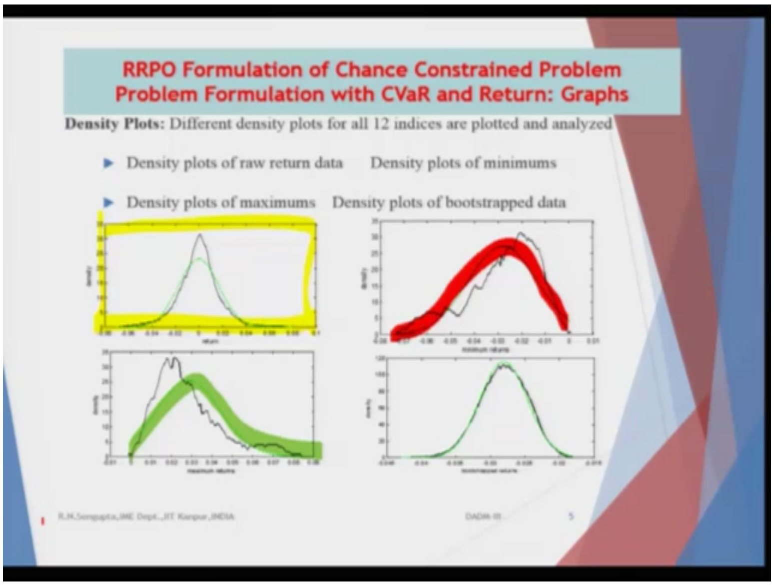


We take, as I considered and as I mentioned that we take 12 indices from the world market, they are basically starting from AORD, BVSP, CAC from France, DAX from Germany, FTSE from UK then HENCSEN, I am just reading few of them HENCSEN from Hong Kong, NIKKEE from Japan, NSE from India, NYSE which is New York Stock Exchange and Singapore Stock Exchange.

So when we considered, we basically report the average value, so remember the average value would be the hats because you are taking the sample and from the sample you are trying to estimate. The same concept which we consider in DADM-1, you took the sample statistic, utilize it in the, as the best estimate for the population parameter and proceeded accordingly.

So we will find out the mean value, average over the height value, hat values, for all the 12 indices. Then we found on the Sigma hat which is the standard error, estimated values for all the 12 indices, similarly we do it for model 2 and model 3 depending on how the window shifting, and how the concept of overlapping has been considered.

(Refer Slide Time: 7:40)



So these are some of the density plots, density plots for the minimum, density plots for the maximum, density plots for the bootstrapped data. So density plot for the maximum, minimum is an interesting point which I want to mention. So say for example, you consider the windows, total length is T, you break down into windows of length L and shift them by S. So what is important is this, what I am going to talk now.

For each window, small window of length L you basically note down its maxima and minima. So you basically again shift it, take the next window again find out the maxima, minima continue doing it. Say for example, if you shift such windows 30 number of times you will have such 30 minimum and 30 maximum values.

These minimum values and the maximum values are the data point based on which you will now bootstrap and do the simulation run in order to basically find out what are the weights for the portfolio which are being formulated considering that you are you are only concentrating on the minimum, returns of each and every stock or else you are going to consider only the maximum returns for each and every stock.

Now this minima and maxima are coming into the picture but because we are going to consider either the loss distribution or the profit distribution depending on how you want to basically plan your portfolio because your main task is to have an analysis of the risks for a portfolio and a higher risk or a lower risk depending on the circumstances will give you a good idea that how will basically try to minimize the overall risk for the portfolio which you are trying to formulate.

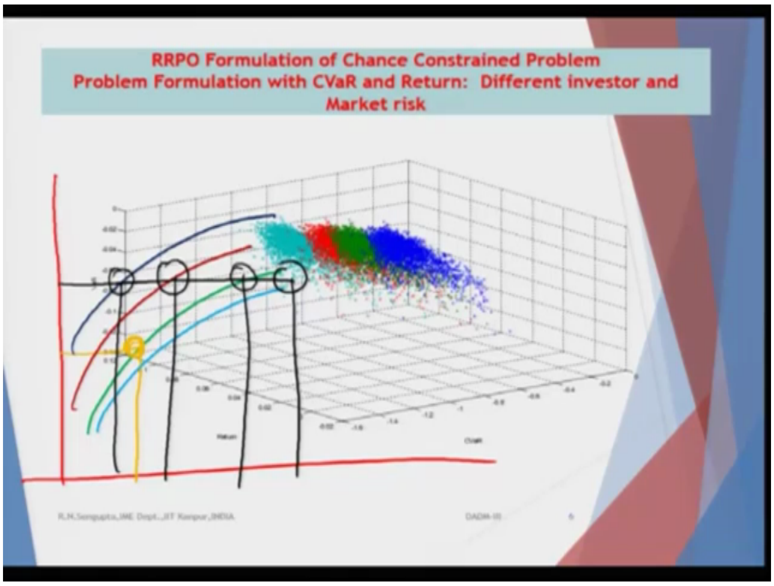
So the first figure which is here which I will highlight with yellow, this is the density plot for the raw return data, just raw return data. You do not take the maxima, you do not take the minima you just take the data. Find out the length of each and every window, find and then take the average whatever is the average it is and then keep shifting the windows and then find out the average at each and every point and then you basically plot the density plots for the return data as such, so that would come out to be a normal distribution considering the central limit theorem to be true.

Now in the other case which I mentioned just few minutes back, you would basically take the windows and keep shifting the windows, take the maxima or take the minima and then utilizing those maxima separately and those minima separately you plot up the graphs, so you basically have one of the distribution. I will just utilize the color, so this is, see the distribution is basically skewed and goes a little bit less steeply on through the left hand side.

In the other case, you have a distribution which basically goes more, not smoothly goes less deeply on the right hand side. So it will basically be either the maxima and the minima which have been plotted. And once you have the maxima and the minima considering that the number of data points is less. You can basically bootstrap it to find out the results of the average value in the mean and the standard deviation for the maxima and minima then proceed accordingly.

So the again remember one thing if the number of data points for the maxima and minima are too large then obviously against central limit theorem would definitely take its overall, have a pronounced effect in the overall data which based on which you are going to work.

(Refer Slide Time: 11:16)



So these are some of the formulations of the and the results for the chance constraint problems considering one of the risk as conditional value at risk which you have discussed in few one other lectures and one is the returns. So different investors and market risk are considered so as I said the investor has a certain perception of risk, the market has a certain perception of risk. What is important to note that you will basically combine them in such a way, that both the market risk as well as the investor risk would be basically portrayed in the analysis.

So if you remember we have discussed that value at risk is basically the risk, the alpha value not the alpha value you do not confuse this alpha value with the alpha value which you considered in the standard normal deviate. This alpha value is the risk basically based on the fact that you are trying to formulate your value at risk, so this alpha value will be taken as the individuals investors risk while the beta value if you remember it was probability of some value greater than equal to beta that beta value would be the value based on which you will try to find out that what is the market risk.

So if I consider this graph, so obviously the beta value is changing, the alpha value is changing, if I consider this graphs and if I take the 2 dimension one, let me draw the two dimensional one first in this graph only and then I will erase that and come to this diagram which is discussed. So you have, now let me change, so this is the axis I consider the distribution as such, so this is the deterministic case that means we are considering the returns as given and the standard deviations as found out and we plot them.

S obviously, if we have we just go vertically up, go horizontally to the left and this point which is being marked would basically be one particular point which will have a particular weight for those particular stocks which have been utilized to formulate the risk return profile and that weights would be very unique to that point where we have just found out what is the risk and the return.

So keep changing the risk and return if they are optimal you will basically be able to draw the whole curve which is green in color, this curve. Now the interesting part is that the returns of the stocks are (themselves) by themselves non-deterministic, so if they are non-deterministic obviously, you will have different frontiers which are almost parallel to each other. So this would basically the blue one, the green one and the and the darkish red one would basically be the different type of risk return profile over, remember one thing I am trying to basically draw the risk and return profile on the Cartesian coordinate.

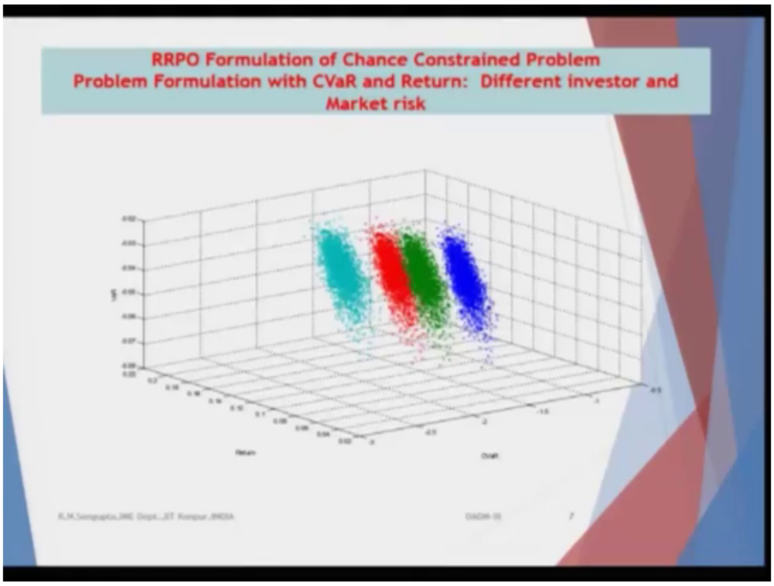
So these three graphs would give you the risk return profile corresponding to the fact that you are trying to change the probability, that is the probability is the beta value based on which you have formulated the problem which is basically corresponding to the market risk which you have. Now if you look at the graph which is shown here the blue, the green, the red and the greenish blue. So if you look from this line the graphs are exactly same as I have drawn that means there are 4 graphs and these 4 graphs are almost overlapping each other, they are concentric, so called not concentric circles but they are like one top of the other almost parallel.

So if each of them are parallel if I am looking from this side the risk and return is here. So if I plot it, so I will basically have 4 such contours one of the other, if I and this is what I have done in, drawn in this graph which is shown on the slide, if I go vertically up, I will cut this or touch these 4 contours at different points and if I then basically go horizontal left, my risk and return profile would basically be given by a different level of risk or return depending on how I have been able to draw the graph.

Whether I am trying to draw the return on the y-axis or I am trying to basically draw the risk on the y-axis. Similarly if I go from the right hand side that means from the y-axis and I go horizontally onto the right. So obviously in that case so let me draw, so I will basically touch these 4 contours on this overlapping curves at 4 different points. So these 4 different points when I drop them I will find out they have different level of risk.

So different levels of this would basically be portrayed by the fact that what is the level of confidence which I have from the answer which is basically how robust or how reliable my answers are.

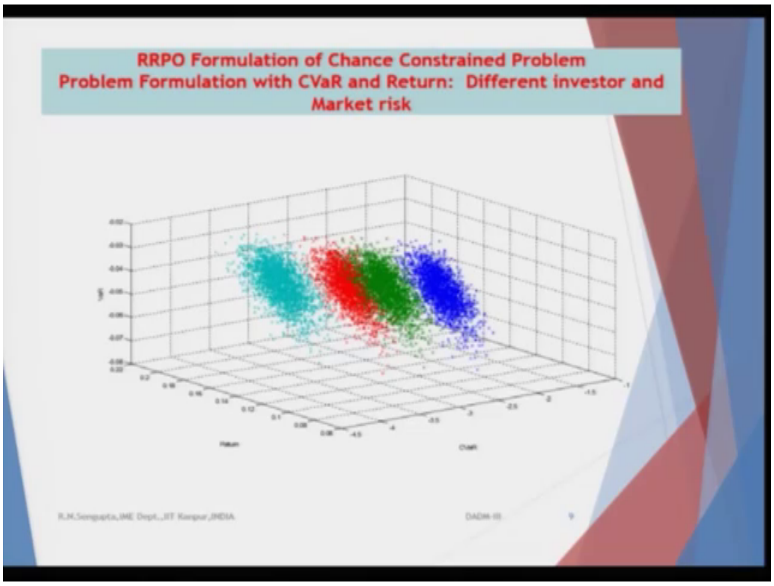
(Refer Slide Time: 16:28)



Then here I show different figures, so in this different figures or what I am trying to plot I am, okay by the way remember I did not mention that. I am trying to plot the VAR value along the y axis and along this axis I am trying to basically plot the return and the C-VAR which is another form of risk I am plotting on which is not orthogonal to the return, it is orthogonal to the value at risk also but is orthogonal to the returned considering the base.

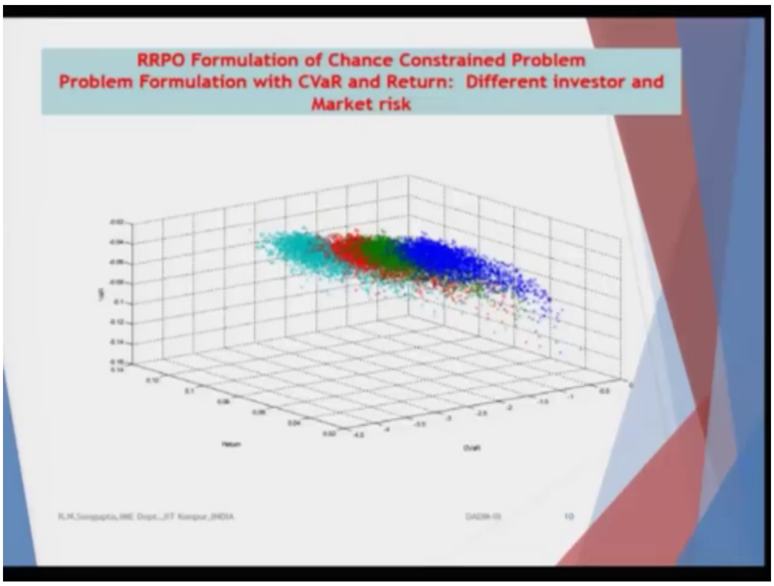
Again I will have definitely different contours, contours one overlapping the other and then basically trying to draw a vertically up or draw horizontally on to the right would give you different levels of risk and return pertaining to the beta value which I want for my answer. This is another set of graphs problem formulation with C-VAR and return and VAR, different investor and market reserving are being portrayed, this market risk would come basically from the value of beta and individual risk will basically come from the value of alpha which we have already discussed.

(Refer Slide Time: 17:27)



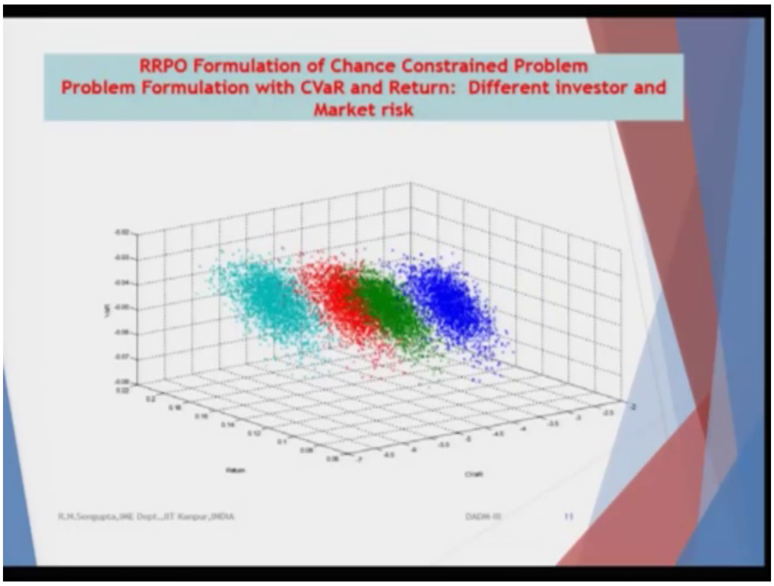
So these are different combinations, so if you remember we can take different combinations for the fact, that for beta we will have 4 that means we can consider beta 0.9, 0.92, 0.95, 0.97, 0. 99. Similarly the value of alpha can be also be considered as 0.9, 0.91, 0.92, (0. 97), 0.95, 0.97, 0.99. So any combination, so you have basically 7 different combinations for the beta and 7 different combinations for alpha, you can basically plot a combination accordingly. That means for each a value of alpha you have so many numbers of betas and so on and so forth. So this I have I am just telling you I am not drawn all the graphs accordingly.

(Refer Slide Time: 18:13)



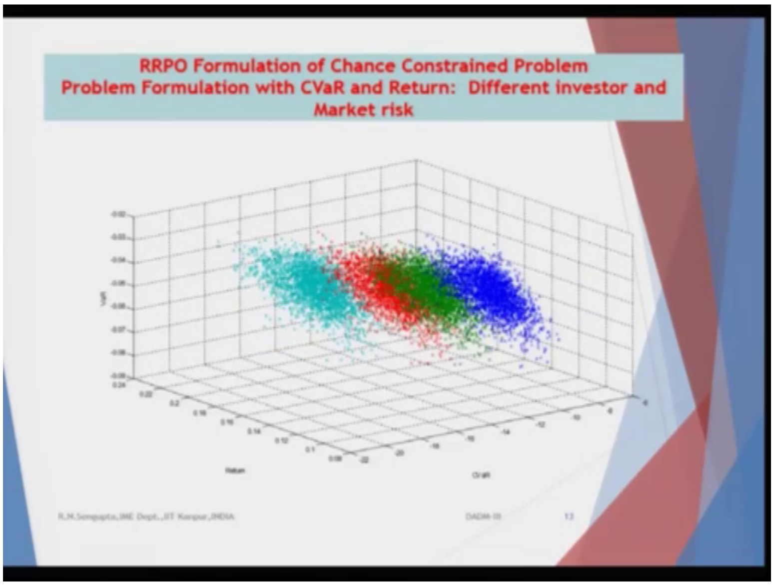
The problem formulation with C-VAR and return, so different investors and market risk basically plot the risk, considering the values of alphas and betas they are again drawn and you can find it out here.

(Refer Slide Time: 18:27)



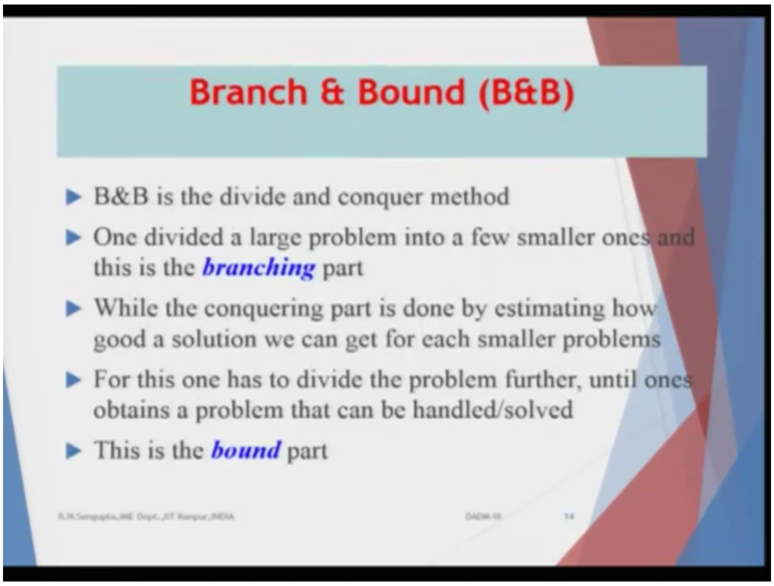
This is another set for different values of alpha and beta, I am just showing you how the graphs look, characteristics wise they look the same, that means they may be shifting here and there but they move in such a way that they are almost overlapping to each other so drawing a vertically up or going right horizontally will give you different combinations of the risk and return profile for different values of beta and alpha. This is another combination where I find out the risk return profile for a different value of alpha and beta.

(Refer Slide Time: 19:03)



Finally so, this is another combination wherever I see that I have the risk return profile, where risk return profile considering the risk word is being utilized very generically where I use the word bar also or use the formulation of bar also to find out the risk and I use the concept of C bar also to find out the risk.

(Refer Slide Time: 19:24)

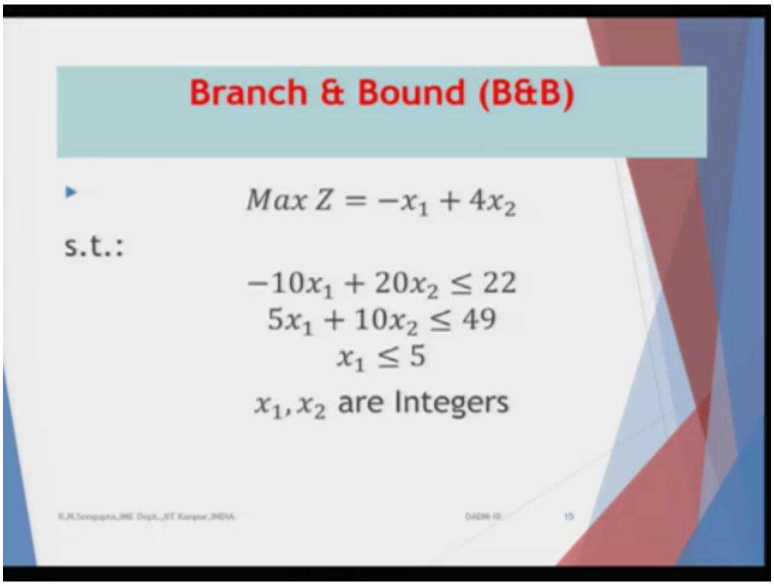


Okay, now we will basically consider the concept of branch and bound. So the branch and bound concept would be considered in a very simple example. So here the branch and bound technique if you remember we have said that one and divides a large problem into a very similar set of problems and they would be knowing on the branching.

While the conquering part is done by estimating how good a solution can get for each small problems or the change of the problems which you get and we try to basically verify how good or bad the results are. For this one has to divide the problem further that means you go stage by the stage problem 1 you branch and branch out in 2 directions, then in stage 2 again you branch out in 2 direction but you bound them depending on whether you are able to get feasible solution or higher than feasible solution which is actually required or infeasible solution, obviously you would not proceed in that direction.

So for this one has to divide the problem further, until one obtains a problem that can be handled and solved, so this is basically known as the bound part, so branching part you branch out go in two different direction, bound means you want to basically take a decision based on the result you want to stop in in 1 direction where you do not want to proceed.

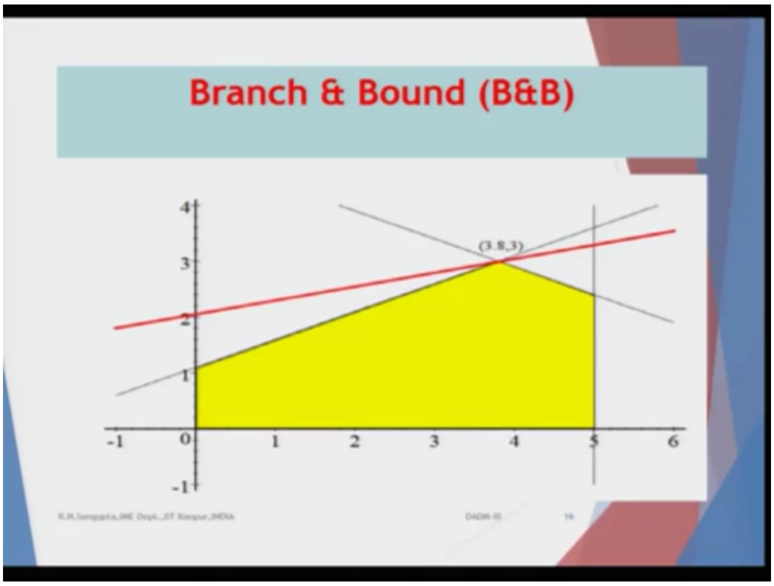
(Refer Slide Time: 20:46)



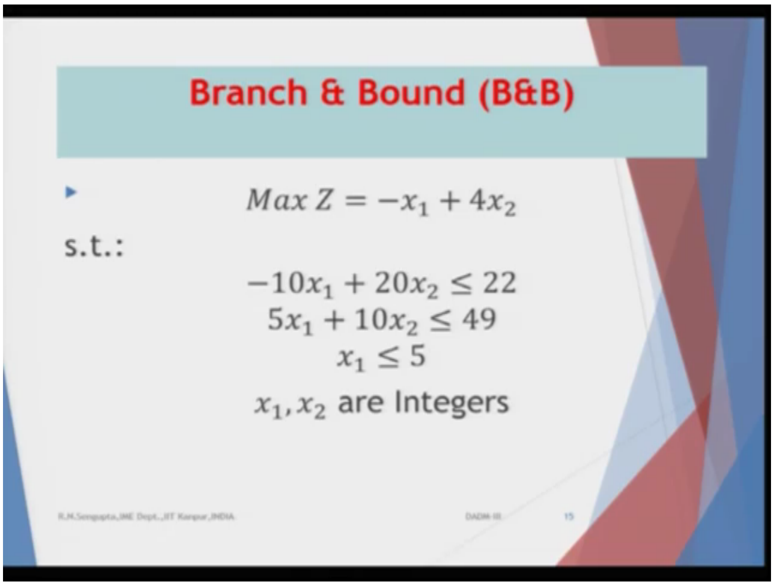
So let us consider this simple maximization problem which is minus x1 plus 4 x2 and the constraints are minus 10 x1 plus 20 x2 is less than equal to 22 and 5 x1 plus 10 x2 is less than equal to 49 and your values of x 1 is less than 5 and x1 and x2 are all integers. So obviously the values of x 1 if there are less than 5 it can be 0, 1, 2, 3, 4 and 5 also included because it is less then equal to 5. While the value of x2 can be 0, 1, 2, 3, 4 till whatever values it is feasible or possible.

Now when you see this problem it is you have to basically first check as you remember we had basically first to formulate the last LP relaxation, solve it, check the answers whether they are integers, check the objective function also if they are not integers then you basically change the values according to the concept of branch and bound and then basically proceed accordingly.

(Refer Slide Time: 21:52)



(Refer Slide Time: 22:04)



So this is the graph, so this is basically the whole yellow area is basically the feasible region, so let us check the first one, so this one x is less than 5 is this. So if I consider this 5 x1 plus 10 x2 is less than equal to 49. So any value of x1 is 0, x 2 would be 4.9, x1 being 0, x2 would be 4.9, so you will find out the values so this is the value of (4 point). So you can find out the value of the graph, so 1 line would be constraint would be constraint would be this one, another line would be this one, so your actual point is this which has a value of 3.8 and 3.

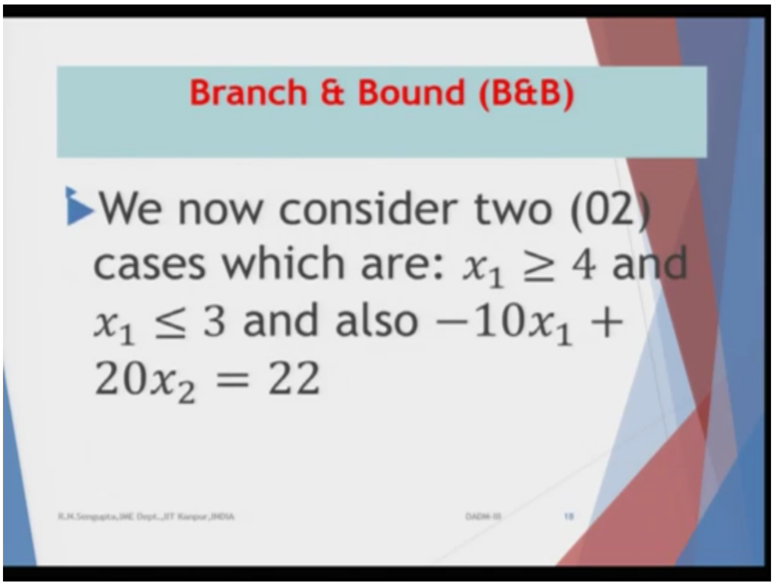
Now coming back to the objective function which is interesting, minus x 1 plus 4 x2, so if your red line moves towards, so obviously you are trying to basically go for the minima but it moves out, out, out and at one point of time it leaves that overall feasible region and that technically should be the maxima considering that you are going to take the linear programming formulation.

But check whether the answers are integers they are not so obviously you have to relax them and considering the branch and bound we will basically proceed in the direction where the problem, change in the results depending on the maximization or the minimization problem happens in the right direction.

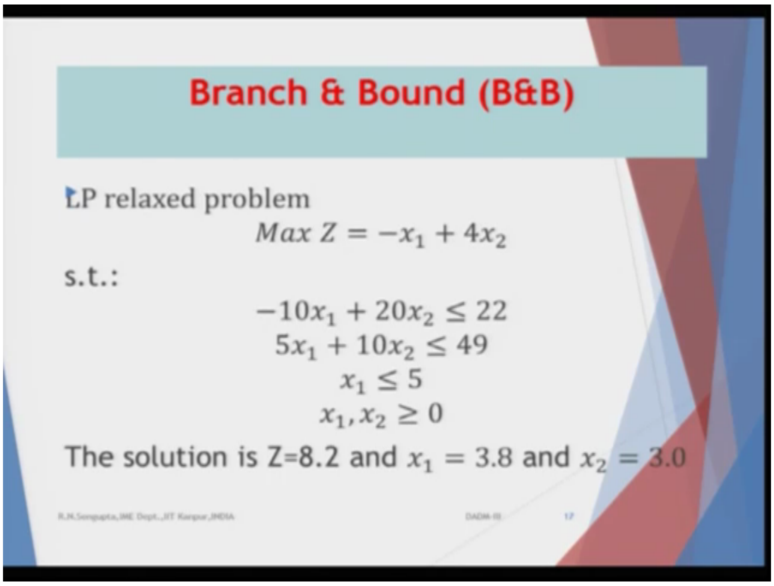
Now you basically relax the problem and solve it, for solving you are using the linear problem formulation you max find out the max value which is again this problem formulation, objective function remains same the such then conditions is minus 10 x1 plus 20 x2 is less than equal to 22 and the other part is basically 5 x1 plus 10 x2 is less than equal to 49 and the value which you have because you have considered that the x values are x1 is less than 5, we will consider that x1 value as your relaxed you are going to basically consider x 1 as all the values which are possible between 0 and 5, point 1.

Point number 2, you have considered x 2 as greater than equal to 0, but initially it was integers but now if you are basically relaxing it will basically be all the values starting from 0 to infinity. So when you solve it as I said the Z value comes out to be 8.2, x 1 is 3.8, x 2 is 3. Does x 2 satisfy it is an integer? Yes it does. Does the value of x 1 which is 3.8 satisfy? It is not, so basically you have to go into the concept of branch and bound.

(Refer Slide Time: 24:53)



(Refer Slide Time: 24:59)

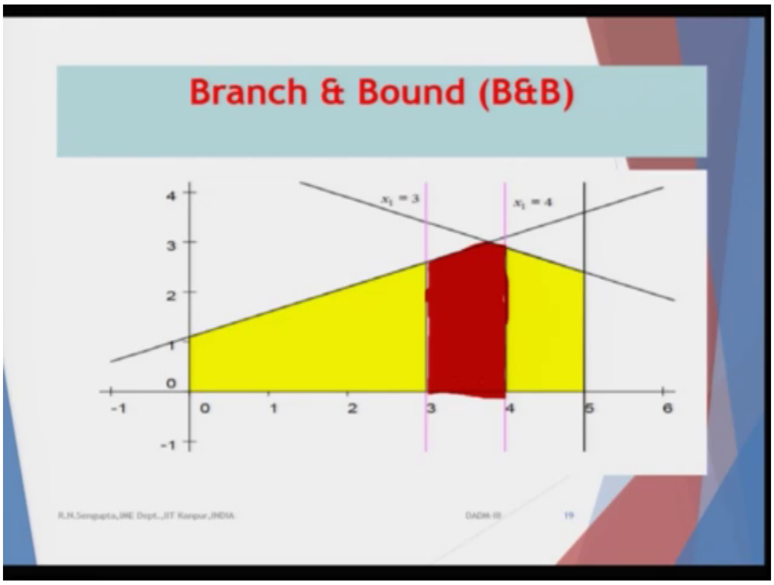


We now consider 2 cases, so check here, the what was the value? Value was 3.8, so we will basically go for values greater than 3.8, the integer 4 and above, so we will divide into that region 4 and above for its x1, for 4 and above, so 4 and (above) value would be applicable and also the 5 value will be applicable because any values greater than 5 are not done or not allowed under the assumptions of this problem.

So on the right hand side you will basically have x 1 is equal to 4 and 5 and the left hand side the values of x 1 would be 0, 1, 2 and 3, so you will basically divide into (3) 2 regions so these are the one. You will have basically x1 is greater than 4 and 5, and x 1 is less than equal to 3, so it will be 0, 1, 2, 3 and also we will consider that the equation which you have means

10 minus 10 x1 plus 20 x2 is less than equal to 22, you will basically put it as an equality sign and try to basic you solve it accordingly to check the feasibility of the solution.

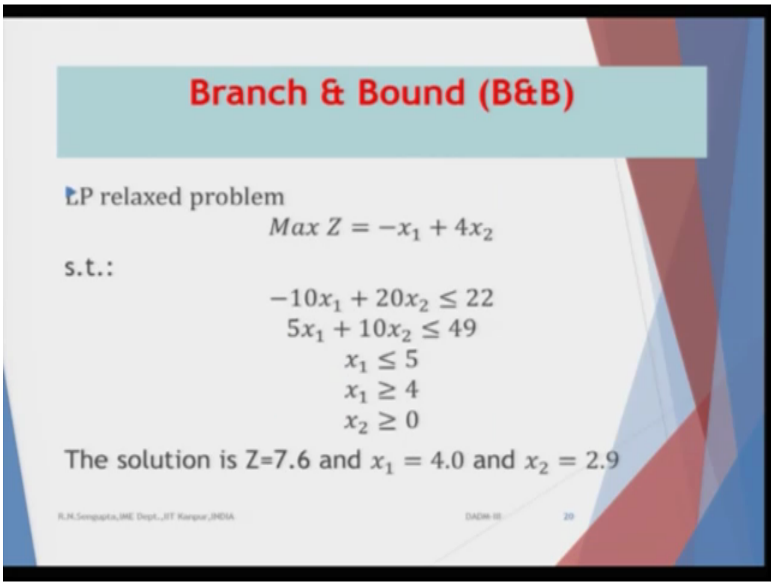
(Refer Slide Time: 26:09)



So this is now the interesting part, this region that I will try to utilize, so this let me draw, wait-wait this is not allowed so this color is red, so this region which was fine according to linear programming problem is no more feasible, no more allowed so this area, whole area would not be allowed, this region is not allowed, so this region is not allowed.

On the left hand side, what is there? We know x 1 with basically would have values of less than equal to 3, so 3 is allowed, 2 is allowed, so there are only integer values, then 1 is allowed then 0 is allowed. But interestingly we will still continue relaxing the problem but basically divide into 2 regions. So in the first case, x 1 is greater than equal to 4 and less than equal to 5, in the 2nd case, x 1 is basically greater than equal to 0 and less than equal to 0.

(Refer Slide Time: 27:34)



So again we formulate, so when we solve it we will basically have to change the values according to what x 2 is coming, so basically consider x 2 as it is keep changing, now you are trying to basically going to branch or consider x2 as such and then proceed accordingly to solve the problem. I will discuss this problem in the next class, have a nice day and thank you very much.