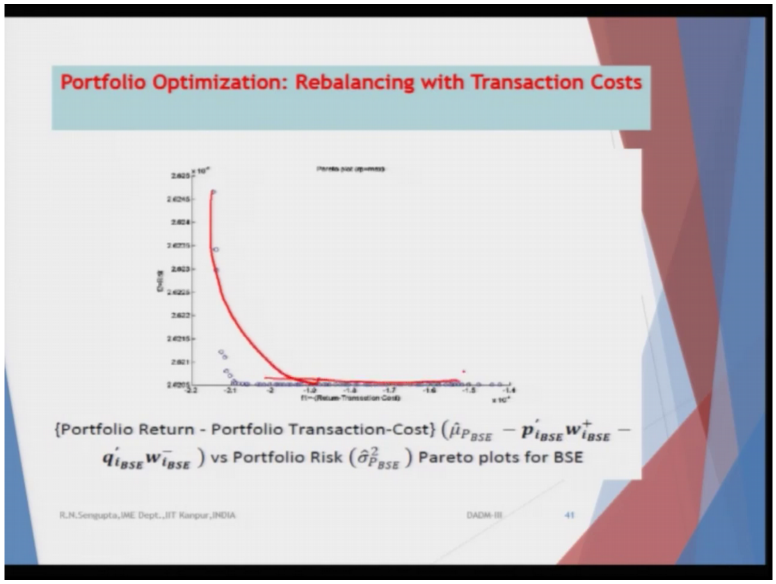
Data Analysis and Decision making-III
Professor Dr. Raghu Nandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur
Lecture 45: Data Analysis and Decision making-III

Welcome back, my dear friends. A very good morning, good afternoon, good evening to all of you wherever you are in this part of the globe. And this is the DADM three, which is data analytics and decision making three course under NPTEL MOOCs series. And this total course duration is spread over 12 weeks, which is 30 contact hours, which when converted into number of lectures is 16 number considering each lecture is for half an hour and each week you have five lectures totaling to each being of half an hour.

So you have two and a half hours of lectures in a week. And after each week you have one assignment. You are, so if you see this, the 45th lecture, which is the last lecture for the ninth week, we have already taken eight assignments. After the ninth weekend you will take the ninth assignments and in totality you will take, you will basically be taking 12 assignments and after which you will basically be appearing for the final examination.

So for the other topics of a objective, multi objective programming, the concept of optimization, goal programming, I was basically going through the problems. So in this 45th week, which is the last lecture in the ninth week, I will discuss the further the analysts of the results and rather than going to the detailed solution techniques, I will come to that later.

(Refer Slide Time: 1:43)

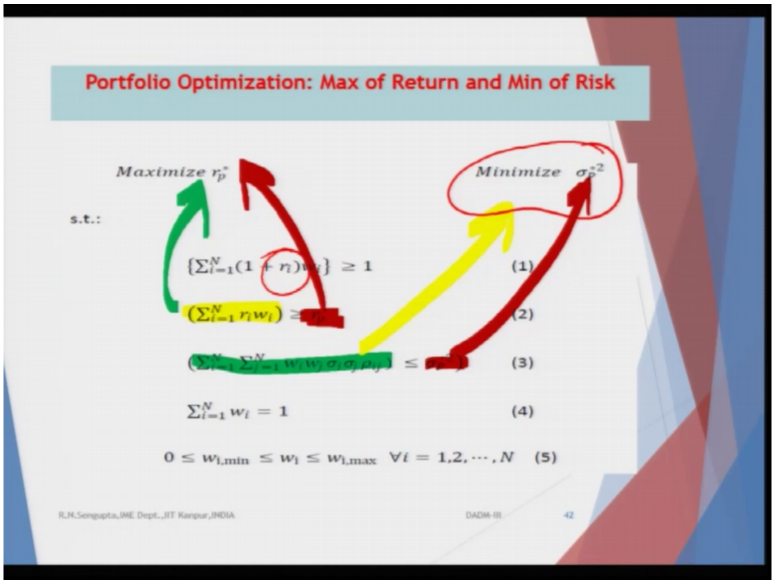


Now if you consider the problem, which you were discussing, where it was trying to and that quantity programming concept, which was first part was half X, I am not talking about the transports, it is X into Q into X, which is exactly the same thing as you did for finding out the variance covariance metric multiplied by the square of the weights and the other part was c into x, which was basically summation of a wi into ri bar.

So here X are their decision variables, which are wi and c inturns, which we are considering. So if you considering the portfolio return corresponding the portfolio transaction cause that means each transaction of buying and selling basically entailed some costs of w plus and w minus, even though both the values are outflows from your pocket but I am just denoting as w plus and w minus with the suffix i corresponding to that Ith stock.

So if I consider the transaction cost for the portrait returns versus the portfolio risk, I would basically have a graph accordingly like this. So these maximum values are towards the X axis, which is very difficult to differentiate here.

(Refer Slide Time: 3:01)



Next we will consider an outline on the portfolio maximization of return and minimization of risk. Now there you can basically combine them using the concept of convex combination function. Now it could have been like Lambda weight between zero and one Lambda in two returns plus one minus Lambda into the risk. So considering that you are putting weightages. Also it can be the case where the Lambda and one minus Lambda or Lambda one Lambda two can be higher values of, higher positive values depending on the total concept of weight or importance you are going to place on those metric whether returns or risk that is a different question.

So here you have maximization of the returns and also minimizing of the risk. Now this part is actually the way how you are going to formulate the quadratic programming, But remember there is a difference here. So I could have minimize, the actual portfolio variance. In that case you would, it would have been six suffix P without the star.

The star means that I am pre defining some value over on which the portfolio risk should not be. That means it has to be less than that or you want to basically bring down the concept of that risk? So this is something to do with those safety first principle, which we are considered in the utility concept where trying to basically shift the mean depending… there is a distribution which is a normal one depending on maximization minimization, we may try to shift the whole distribution mean values either to the right or the left considering the number line.

Or else we can shift or try to basically increase or decrease the value of R suffix L or R suffix F, F is the risk free interest rate and maximization of rather than RP star we can could have basically maximize RP, which was technically the summation of wi into ri bar. Similarly, as I just mentioned, sigma squared P without the suffix actually is basically a double summation wy into wj sigma ij where sigma ij is covariance values.

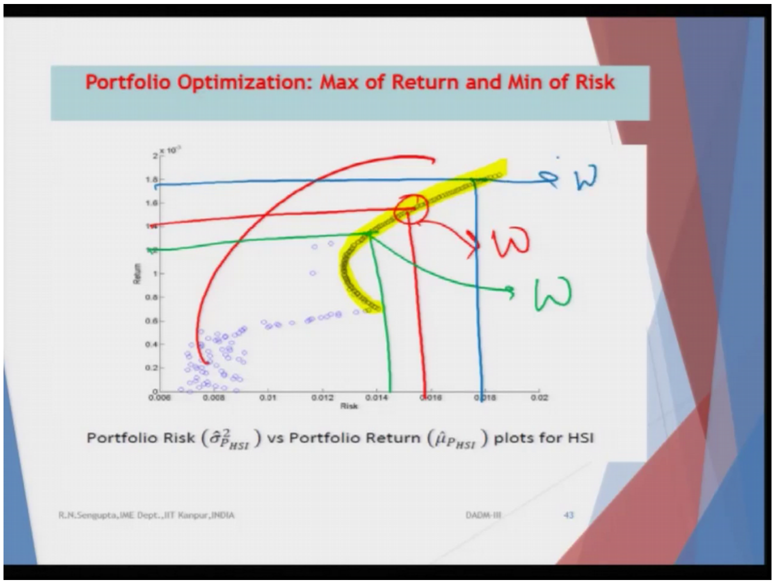
Now the what are the constraints, the constraints are in the simple case that if I have the returns, so if the returns are increasing, there would be some increase and decrease. So the returns can be the positive or negative. What we will consider that in the long run that combined returns of the portfolio values would be greater than what we have actually started with. So some of the values here.

These ri can be negative also. The second one is that portfolio returns. Now actually, which is what I just discussed so this value, the value here actually could have gone here and the value here could have gone here. That means I replaced the actual value with the fixed value, which I am going to increase for the rest case of the returns and trying to decrease for the case of the risks. So these are the values which are being transformed. So this has gone here and this has gone here.

So we will ensure in this case the yellow highlight ones means the portfolio actually returns is greater than RP start as it should be, because I want to ensure the return on the portfolio is over a fixed value, which may be market dependent, may be depending on my uncertainty, maybe depending on my concept of utility, whatever it is. Me means I am the investor.

Similar in the case, the risk would be the whole risk of the portfolio. That is double summation wi into wj, into Sigma i, Sigma G into row ij or Sigma ij would be basically the portfolio risk, which has to be less than equal to Sigma squared star suffix P, which is again a predefined value by the investor or as per the concept of the utility and obviously the summation of the weights are equal to one and each of the weights wi are between wi minimum and wi maximum.

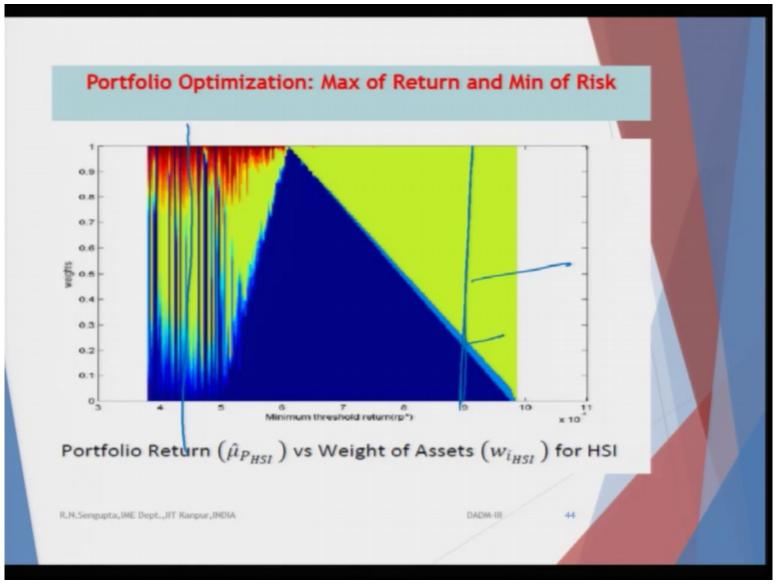
(Refer Slide Time: 7:39)



So if I basically plot the return and the risk, so the return and risk of the as it should be. So this basically mimics, so there are obviously some simulation values which are not matching it, but in generally if we see the trend, so these values are exactly what we want. So this is the risk return profile, Risk along the X axis and return along the Y axis. So if I basically want to find out any particular portfolio, their corresponding weight and returns, I basically find out these are the weights, here w is a vector.

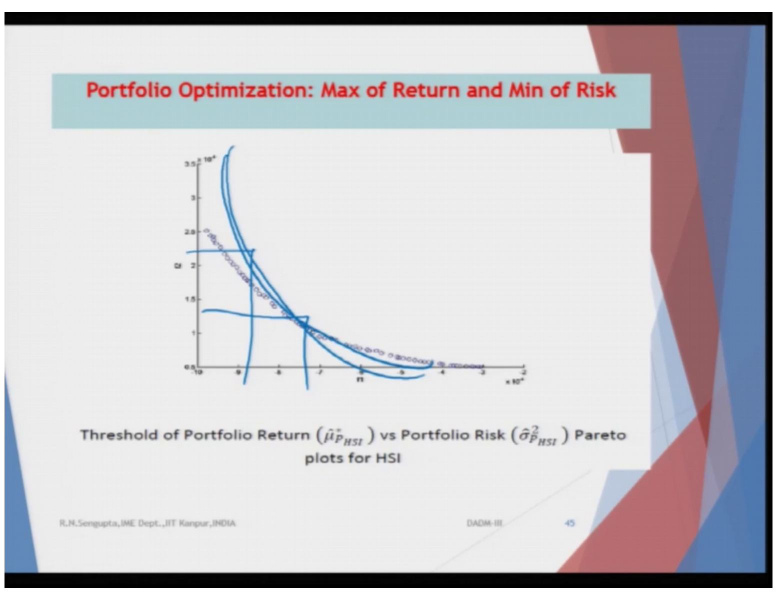
Similarly for other case of risk and return, I would have some other weights corresponding to third value and what all than weight. So these are the different combinations which I have. If I know want to know the weight, obviously I will plot the minimum threshold value, which I have RP star and then the weights are plotted along the Y axis.

(Refer Slide Time: 8:50)



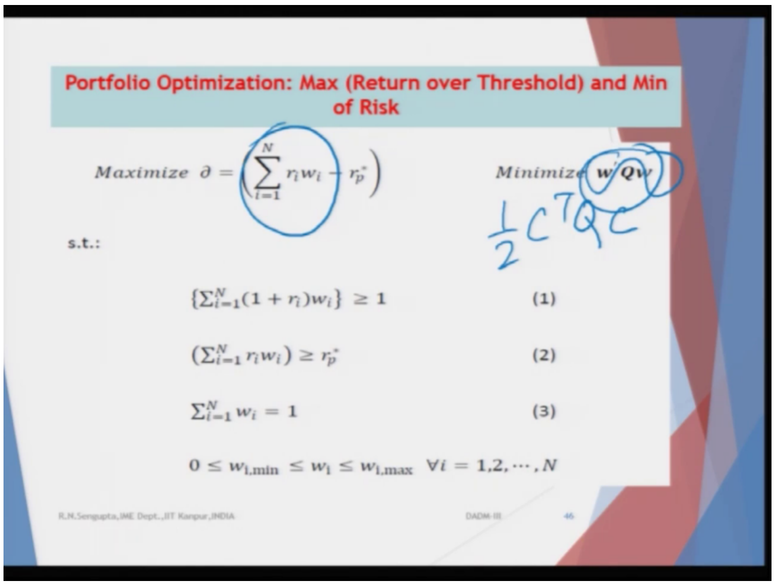
So these at any value threshold value say for example threshold value is nine into ten to the power minus four, So at this instant there are only one stock, the dark blue one light blue one and the green one. If I consider any value here. So here there are different colors combinations. It is very difficult to see, understand, so here the different color combinations. So there are much more number of stocks which are basically being utilized to formulate that portfolio, which is the optimum one.

(Refer Slide Time: 9:31)



Then if I basically have the concept of the threshold of the portfolios, threshold value remember and the portfolio risk, which is the Pareto plot and this I am doing for the Hong Kong stock exchange. So that Pareto plot as it should be, this is the value. So here any combinations would give me that level of utility for any combinations along that line gives me the same level of satisfaction.

(Refer Slide Time: 10:00)



The next value which I consider it is basically in the maximization of the return or threshold value of the return and minimization of the risk. Now, see here the again, the interesting part, this is exactly what we had seen earlier. So this is the quadratic one. So you have half CQ and into C which is basically C squared. C squared is basically the weights which are, the one values which are here. And, and Q is basically the variance covariance matrix.

And I want to basically find out the this is the minimization. Obviously this has to be minimize and maximization would be corresponding to the threshold value. That means there are certain value over on which the portfolio return should be. So the threshold value is RP star, which is fixed by the investor depending on his or her utility and the market risk or the perception of own individual risk.

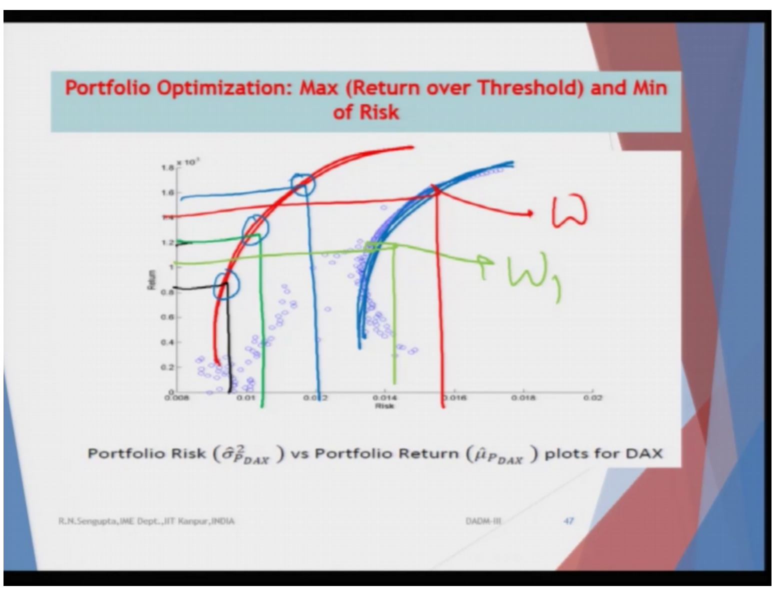
And the return on the portfolio is given by the summation of ri into wi. ri are again, ri or ri bar depending on how you have been able to formulate the problem. And wi are the weights corresponding to the port, the each and every skips you want to find out. So again, the constraints remain the same. I am trying to keep the constraint as simple as possibly.

You can make it more complex, more practical oriented, but these answers are quite well enough give them a lot of information that how you basically should formulate your portfolio. So in the first case, again you are the first constraint basically means that there would be some ri which are positives, there would be some ri which are negative.

But you will basically formulate that in such a way that one plus ri multiplied the weights which you have should be greater than the value of one. One means basically the returns which you are ensuring as when you start at a value of a hundred or one or 2000 whatever it is that has to value, has to be greater than 2000 or 100 whatever.

And again, you need to ensure that the returns of the… you are trying to maximize the difference so obviously the returns of the portfolio should be greater than RP start value as it should be. And the, some of the weights are one actually as it is and the weights are always bounded between the wi mean and wi max for each and every I value I being one to capital N then if I, so I am doing it for the DAXs.

(Refer Slide Time: 12:37)

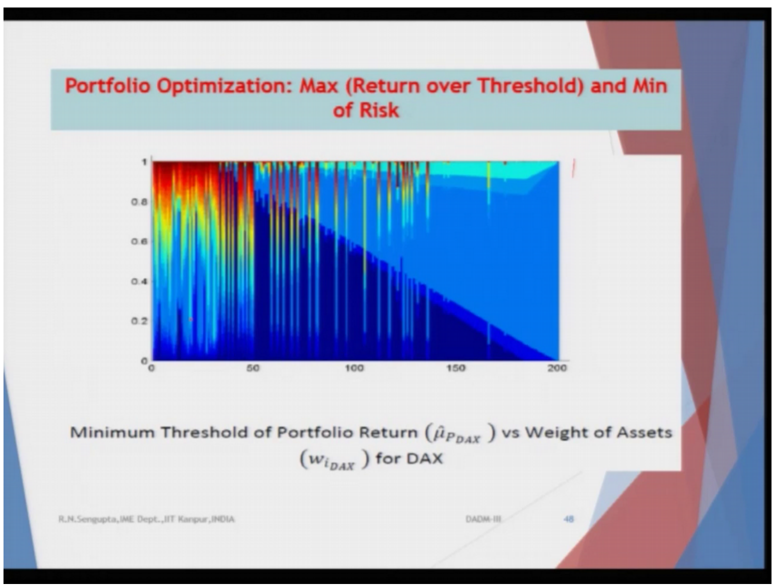


If I basically plot the risk return profile, risk being along the X axis or return being along the Y axis, I have the graph as it is given. So these are the graphs and these are the simulation value.

They are coming out to be quite nice with respect to the actual theoretical values. So theoretical values are this, so I am just drawing them apart in order to make you understand how the shape would look like.

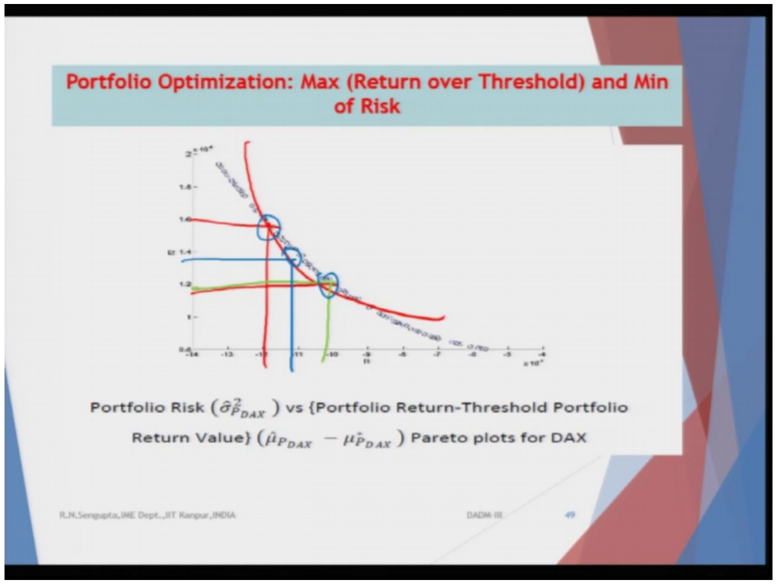
See here again a risk and return profile. Then again, risk return profile value. Then again, risk return profile value who gave me different weights. In the similar way when I find it out, risk return profile gives me w 1. For that 42 instant. Then again, if I basically plot it I get another w value. These are vectors, remember. So similarly I keep changing the threshold values and I basically can simulate such hundreds of points in order to basically find out how the graph will proceed.

(Refer Slide Time: 13:51)



Then again, when I, depending on the number of simulation runs, which I do, I plot the values of the weights along the Y axis and the minimum threshold of portfolio returns, which is for the DAX case, the X axis. Now here you will see that I am basically changing the different in this in order to make you understand that the for the same type of equations, they would give you very similar type of interesting results for different type of stocks, which means that the models which you are trying to utilize are quite efficient enough, decently efficient enough to give you interesting results. So you may find out that as for the higher values of threshold, see for example, only one stock dominance, others are gone. But if you basically go for lower values, you see there is a huge amount of variations in the weights of the total portfolio. So that means there are different types of scripts.

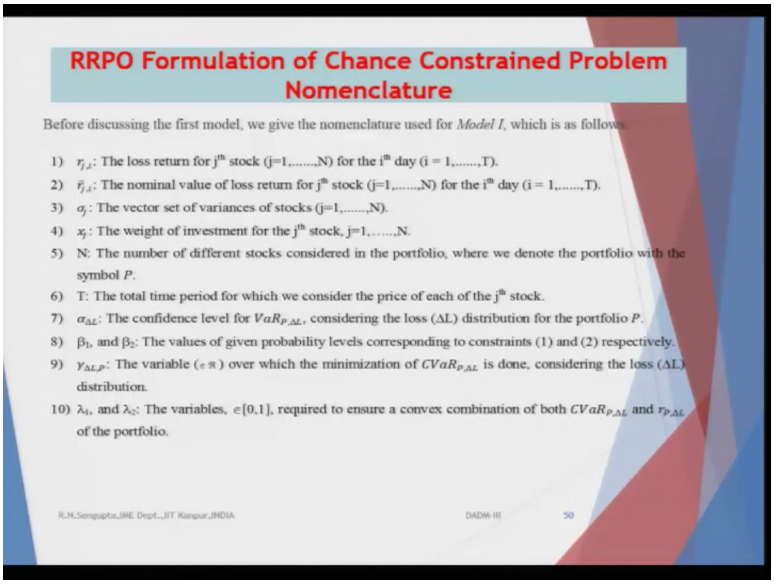
(Refer Slide Time: 14:50)



Again, when I do it for the DAX, trying to basically formulate the Pareto optimality frontier for the portfolio risk with the value of the portrait returns, return means over the threshold values.

So again, I get this Pareto optimal front, so these values, so they will give me the off that Pareto optimal front, the level of satisfaction would be same for all the points but how you are able to combine the different weights is basically what is the question.

(Refer Slide Time: 15:35)



Now we will just consider another simple formulation. I will come to the RRP formulation later on for the chance constraint problem nomenclature. That is what we will do when we consider goal programming. So just to give you the idea, so you have the returns of the jth stock. Again, j is equal to one to capital N for the ith day. So technically when you are trying to basically formulate with respect of time, time should be one nomenclature by t but we have taken here as i. Here r by ji basically a nominal value of the lost return of the jth stock at the ith day.

So you will basically find out, try to find out the average Sigma x and xj are the corresponding vector of variants of the stocks and the weight of the investment for the jth stock, which is of interest to us. And N the number of capita N is the number of stocks considered in the portfolio. T is the total time period, like 240 days, one year or 480 days, two years.

These values of alpha and beta, which is very interesting is basically there would be two levels of risk. One is what the investor is perceiving from his or her side and there is a market risk. So it may be possible that I think the market is very risky but actually it is not. So my perception of value of alpha based on which I am trying to basically formulate the problem.

And the value on the market is in in reality which is beta maybe different so alpha can be greater than beta, alpha can be less than beta. The values of alpha, beta if they are very close to each other, that means I am able to, formulate my problem with the set of information which I have from my side basically mimics or tries to basically be as close as possible to reality and find out the concept of market risk.

Here I am denoting alpha as some level of confidence for VaR. What is VaR I will come to that later and beta is basically the value of given properties corresponding the constraints one and two. So obviously there would be some constraints which would basically imply that what is the level of such risk or level of confidence. I am going to put two on onto this constraints depending on the market values of market values fluctuations of the actual information set or the data set.

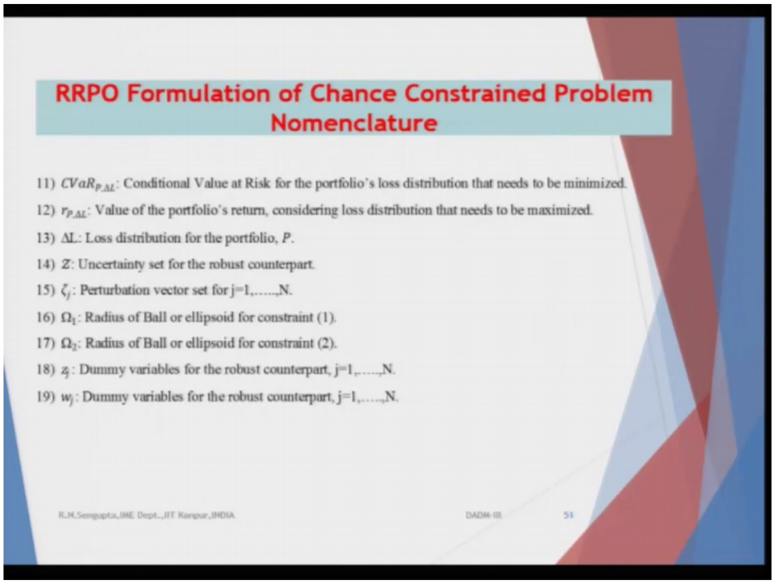
Why? Because the prices of the stocks are for fluctuating, so if they are fluctuating it would have an effect on the returns which is R, and if the returns are fluctuating, obviously they would be some probability based on which not the probability, distribution I am talking about, there would be some a reliability value or robust value between which the value of the stock prices with fluctuate.

And Lambda one, Lambda two are the variables called immune, which basically ensures a convex combination you are going to put for the risk and return and gamma is the variability over with the minimization of the C bar would be done. So alpha is basically for the VaR and and Gamma is basically for the value of CVaR .

So value at risk VaR or conditional value at risk CVaR are the concept which are used in finance to basically find out the overall risk of the portfolio. As we have standard deviation and as we have beta not this beta what is shown here. Beta as we seen, in the concept of of capital asset pricing model.

So these values of as I was talking about this condition, valid risk and value are the risk concept which are used in portfolio as you have as I just mentioned, standard division, also variance also. But remember there are some very unique properties of conditional value at risk CVaR and it has an advantage of the value at risk. This is just I mentioning for your own information.

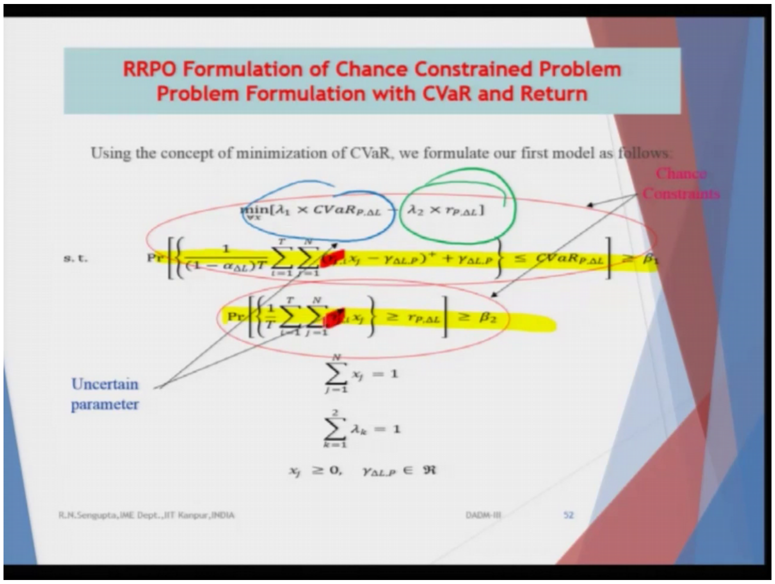
(Refer Slide Time: 19:27)



The other nomenclature or other symbols which are important for us is the conditional value at risk for the portfolio based on the loss. Delta L is basically the loss. I am only considering the loss. It could have been basically for the profit also in, but in that case risk would be given a different picture based on which you, I will try to find out that how good or bad my total portfolio is doing considering I am looking at the positive side. Here in this case I am trying to find out the conditional valued risk of the distribution which is pertaining to the loss distribution only. RPL is basically the value of the portfolio return considering a loss distribution that needs to be maximized.

Delta L is the loss function and the values which is given from 14th to the 18th one are certain to do with type of problem formulation which we will see later and it is 14th to sorry 19th. And do not confuse the, the variable wj as the weights one. It is basically a dummy variable which will be utilized to formulate the problem in the robust formulation being important for us.

(Refer Slide Time: 20:44)



So what we do is that you use the concept of minimization of CVaR or condition value at risk. And also we tried to basically find out the maximization of the return on the portfolio. So minimization of the minus value basically means the maximization. So the first part which we have is the conditional value at risk, which needs to be brought down.

The second part with the minus sign minimization of minus is basically increasing the return on the portfolio and very simply the constraints that have been formulated in a very simple manner. In the first case I am insuring in the probabilistic case, so obviously it need not be probabilistic and the simple stand deterministic also.

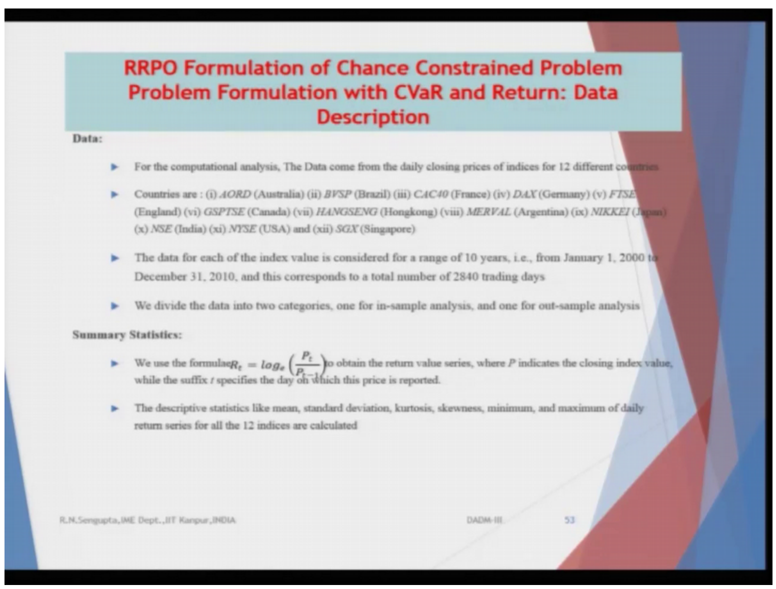
In the probabilistic case it is the property of the actual formulation of the portfolios, condition value at the risk is always less than the conditional value of risk which we are proposing based on our, our level of risk. So that should be in the probabilistic case that if that value is less always then obviously we are 100 percent sure, but if it is not less than CVaR always so there would be, have a level of reliability that is given by beta one.

And in the second constraint you have the, basically the probability of the return of the portfolio is greater than the return which is stipulated, which is RP del L corresponding to the loss distribution. And that the level of confidence is basically beta two depending on how reliable the results are. And obviously it means the sum of the weights of the portfolio of the scripts are greater than one.

And some of the convex formation of the weights lamda one, lamda two obviously should be one. So here the values, which means the uncertain parameters and the John's constraints are highlighted. So the uncertain parameters, John's concerns would be corresponding to the constraints, the problematic sense. And the uncertain parameters actually are the returns only because based on the prices, you have the returns.

The returns being probabilistic you will basically have different type of average returns, different standard division, so on and so forth. So they will keep changing. So actually information based on the prices would basically dictate how the formulation and the values have been obtained. So again, the John's concerns are given, I will just highlight it and the uncertain parameters I have already discussed are basically the R values. So these are returns of the the scripts.

(Refer Slide Time: 23:42)



So the data we are going take is taken from the 12 different countries, which is the Australian exchange, Brazilian, France, Dax for Germany, FTSE for England, Canada, Hansen is for Hong Kong, Argentina, Nikkei for Japan, NSC, not the BAC. NSE for India, New York stock exchange for USA and for Singapore.

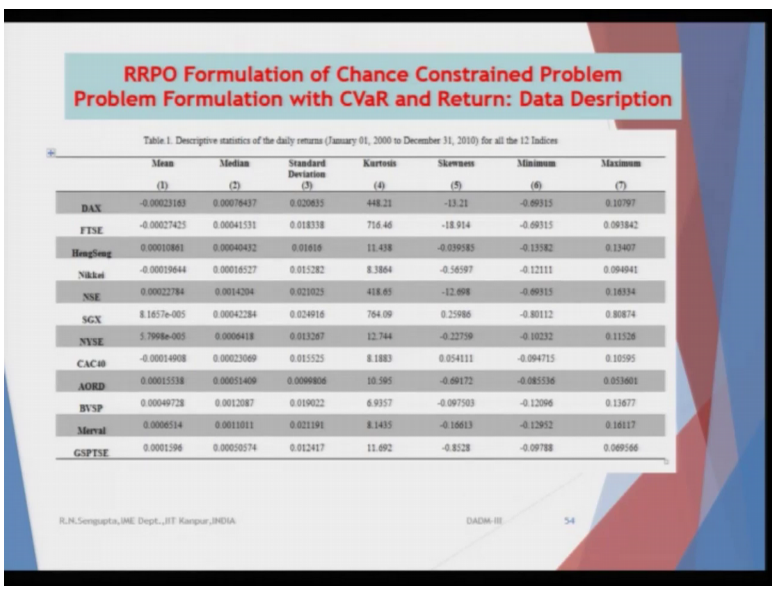
The data for each of the index in value, that means not individually, but for the index which we seen mimics the market to the best possible extent is considered for the range of 10 years from 2000 to 2010 all the 10 years. And this corresponds to a total of about 2,800 trading. Which is two eight four zero trading days.

We divide the data into two categories. So this is important to note that the reason why we divide that data into two regions is basically we want to basically have an in sample and out sample. In sample is where we have the model tested, finals efficacy, and then basically retest the model for data set, which is totally new, such that we want to find out how good or bad our model is or how robust our model is.

We will use the formula of the return. So this capital R and small r are immaterial here. So we will use the, the return on the of the stock is given by LN of P two by P one that is price of day two divided by day one. To obtain the return value series where P closing index price. So you will be taking the prices at the end of each day.

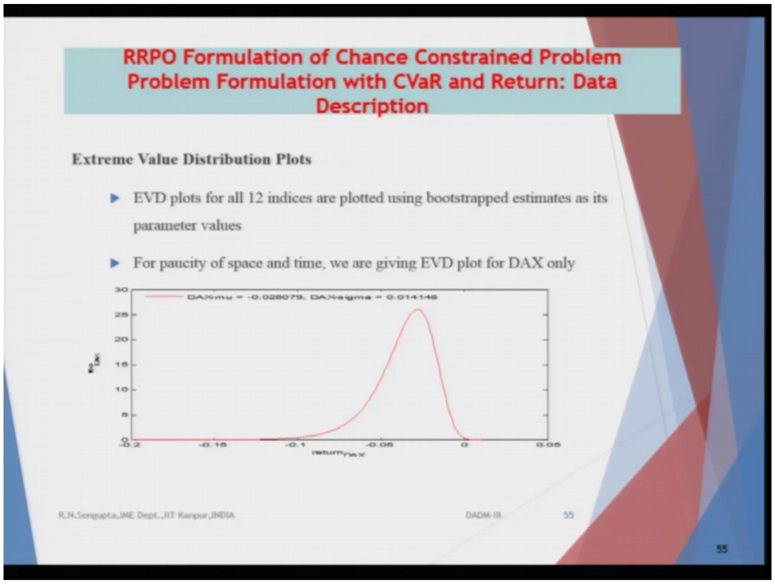
Why the suffix T Specifies the D on which the price is reported? The descriptive statistics where the mean standard division skewness, kertosis, minimum, maximum on the daily returns of all these 12 scripts are given here.

(Refer Slide Time: 25:30)



So on the first column we have all the scripts and the, and the second till the last one are basically the mean values, the median, the standard division, the kertesis, skewness, minimum value and the maximum value. So you can basically utilize the data accordingly and then decide that whether it should be basically a negative distribution, the lost distribution or the positive distribution for the returns.

(Refer Slide Time: 25:59)

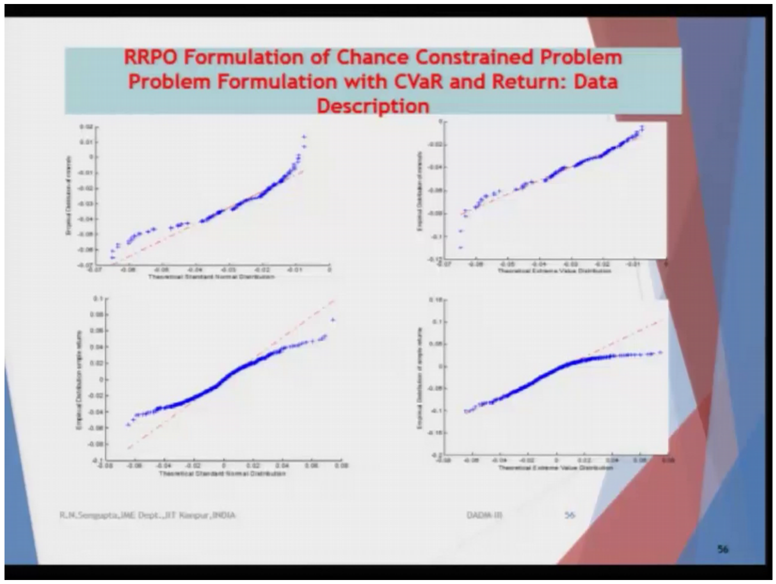


Now we did consider this concept of extreme value distribution in the DADM one. So you will basically, in a very simple manner, so EVDs or extreme value distribution plots for all the twelve indices are plotted using the bootstrap estimates.

That means you find out the estimates for the shifting window depending on how you have been able to formulate and then bootstrap it to basically have a whole set of data runs. And for the positives, we will only give the EVD runs for the Dax only. So this is basically left skewed or right skewed depending on whether you are going to consider the positive or the negative returns of the stocks.

So you will basically have different types of left and right skewed for all the 12 indices and if you see the DAX new values are given a minus 0.028 I am only reading the values and DAX Sigma values is 0.01. These are the theoretical extreme value distributions concerning the QQ plots, which is an important factor.

(Refer Slide Time: 27:13)



We have also considered in some not in details, but we have mentioned that in the DADM one where we considered that a given any of the theoretical standard normal distribution is basically compared with the empirical distribution of the minima. Similarly for the Maxima and we do maximum minimum being for the case for the, the negative returns or the positive returns.

Similarly, we do the theoretical extreme distribution values on the X axis and try to basically find out the simple returns and the distribution of the minimum or the maximum in order to find out how they match. So obviously if it is a straight line, it means that the QQ plots are a fitted the best based on which we will consider that empirical values of the destruction for our case of understanding and consider those values for our general glans.

With this, I will end the nine week and consider more about this goal programming and other concept later in the starting from the 10th week. Have a nice day and thank you very much.