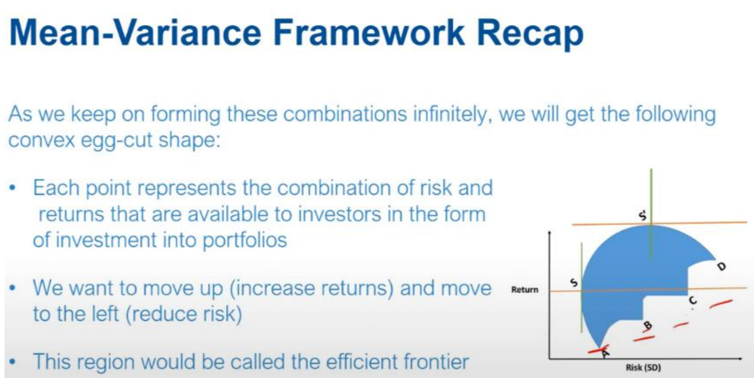
Artificial Intelligence (AI) for Investments
Prof. Abhinava Tripathi
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

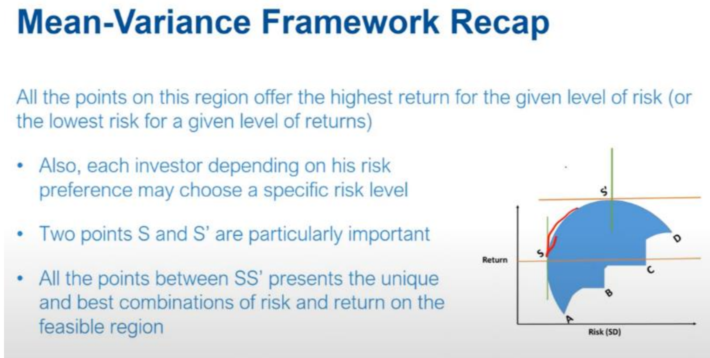
Lecture- 35

In this lesson, we will implement the Harry Markowitz Mean Variance Framework with our programming. First, we will recap the mean variance framework. Subsequently, we will discuss the tail risk measures namely Value at risk i.e. the expected short fall, ES. Next we will download securities data from Yahoo Finance.



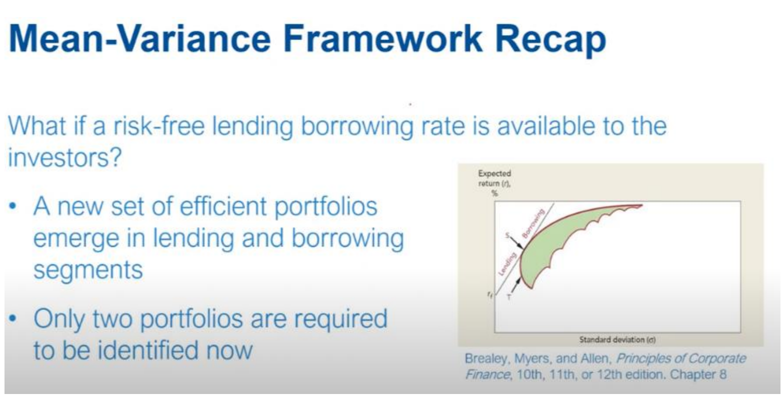
Then we will compute returns and visualize various properties of this data. Next we will initiate a portfolio object, formulate various portfolio constraints including box, group and risk budget constraints. We will formulate and visualize portfolios with different tail risk objectives. This includes equal weighted portfolios, minimum risk portfolio with a target return, tangency portfolio and global minimum risk portfolio.

Next we will construct efficient frontier in an interactive and customized manner and visualize it. We will also initiate a short portfolio object and identify portfolios with different risk return objectives. Lastly, we will also understand how to set up portfolio in Mean CVAR i.e. conditional Value at Risk Framework.



In this video, we will recap our discussion on Mean Variance Framework. Recall the discussion with video topic on portfolio optimization. We started with four securities A, B, C and D and we kept forming different combinations of these securities A, B, C and D and as we keep on doing that in a return and standard deviation space. We obtain what is called feasible region. This blue region is the feasible region which represents all possible combinations of return and risk i.

e. standard deviation. This feasible region is represented by this ecked kind of shape and each point here represents a combination of risk and return that is available to investors in the form of portfolios. Now how do we improve our position in this feasible region? We can either go up or we can move left and as we keep on doing that we reach a particular surface or the surface of this ecked region. This surface of ecked region, this surface is called efficient frontier because every point on this surface represents a unique combination of risk and return which can be explained in the following terms. For example, let us take this point Z.



Now for a given amount of return, this Z would obtain minimum risk or for a given amount of risk, this Z would give you highest return. In a sense, every point on this surface obtains or achieves a very important or we can call best combination of risk and return that is no other portfolio for a given amount of risk can offer you higher return or for a given amount of return can offer you a lower risk and therefore this is called efficient frontier given this property of portfolios lying on this surface. While we cannot directly compare these portfolios on this surface per se, individual investors depending upon their degree of risk tolerance or risk averse nature, they will choose a given particular point on this surface. For example, an investor who is slightly willing to take more risk would probably prefer a position here and therefore he would also be rewarded by a higher level of return. However, in contrast, an investor who is not so risk taking probably more risk averse and would like to take a position here and therefore would be rewarded with this amount of expected return.

Summary

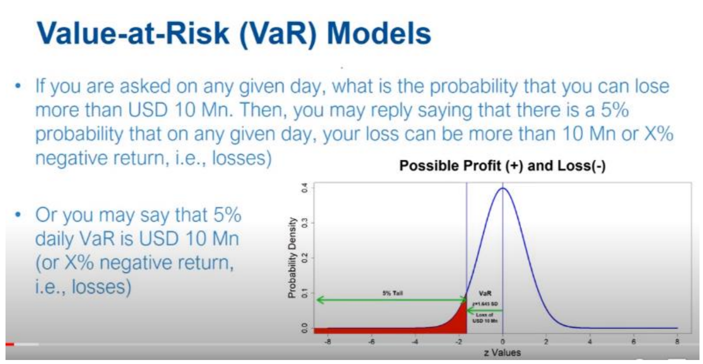
Mean-variance employs variance (or standard deviation) as a risk measure

- There are two more important risk measures that we will implement
- Value at risk (VaR)
- Conditional VaR or expected shortfall (ES) measure

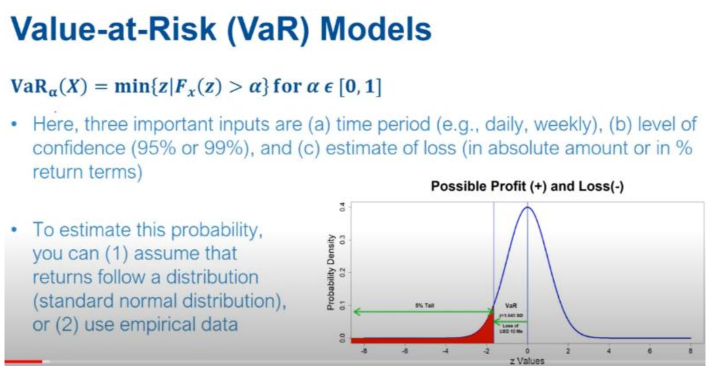
In this backdrop, two points are very important which is one is this S and one is this S dash because this point S offers the lowest amount of risk. This is the lowest amount of risk available across the entire feasible region. So any portfolio in this feasible region does not have lower risk as compared to S. So this point S is also called global minimum variance portfolio because it offers you the lowest variance across all the available feasible portfolios. Please notice this term feasible portfolio.

Feasible portfolios are all the portfolio set that are possible within this feasible region. Another portfolio S dash is important because it offers you the highest return across the feasible region. So any portfolio if you take any portfolio in this feasible region S dash offers you the highest return and the surface this efficient frontier represented by SS dash is a unique and best combination of risk return profile in the feasible region and that is why we also call this efficient frontier this SS dash region. Also while discussing this portfolio optimization scheme and efficient frontier, we noted that if risk-free lending and borrowing rates are available which are same then one can also draw a tangent from RF to this efficient frontier and this tangent line will touch let us say at point S then this S portfolio becomes one of the best optimum portfolio and a new set of efficient portfolios will emerge which comprise the line joining RF to S. We also noted that on the left of this RF segment RF to S this is called lending segment because any point on this portfolio which represents a combination of available risk and return which is RF joining S, RF to S can be obtained by mixing some amount of wealth in RF and remaining amount of wealth in S.

Let us say if you are investing 40% in RF and remaining 60% in S then you can obtain this portfolio segment which is called lending. We also noted that on the right of point S there is a borrowing segment which one can obtain by borrowing at RF because we said RF is available for both borrowing and lending and therefore a slightly more risk-taking investor who is less risk averse can borrow at RF and invest his own wealth at RF at this S portfolio or the most optimum tangency portfolio what we call and remaining borrowed amount as well into S to obtain a position on this borrowing segment and that is how this new efficient frontier is arrived. Importantly this is also called two portfolio theorem or separation theorem because now we need to only identify two portfolios that is RF and the tangency portfolio S to offer a wide range of risk return profiles or what we call the new efficient frontier represented by this straight line. Also now irrespective of investors or our clients risk profile whether they are slightly more risk taking risk less risk averse or more risk averse we only need these two portfolios RF and S to offer them a combination or a choice or portfolio of their taste and preferences that is desirable to them. So we have separated the decision of this portfolio RF and S from the investors risk preference that is why it is also called separation theorem or two fund theorem.



To summarize we recap our discussion on efficient frontier and mean variance framework. We also noted that mean variance framework implies variance or standard deviation as a risk measure. However there are two very important risk measures that we also employ one is called value at risk and another is called conditional war or conditional value at risk or often referred to as expected shortfall measure that we are going to discuss in the next two videos. In this video we will discuss a very important measure of tail risk that is value at risk based models. If you are asked on any given day what is the probability that you can lose more than 10 million dollars then as a first step you will try to plot your returns from negative returns to positive returns these are your losses or these are your gains and then also on the vertical axis you have frequency or probability density of returns.



Often it is said that this kind of graph can be very nicely represented by normal or standard normal distribution. If you are using standard normal distribution then on x axis you have what we call z values or units in terms of standard deviations. This is a very effective way of representing the probability density distribution of variable like return. Let us say you plot this kind of graph and you find that at 5 percent cut off at 5 percent cut off you start from this end to this end and you find that 5 percent area stops or cuts off at a value of 10 million loss here. This point represents 10 million dollar loss and the area under the curve is 5 percent.

So 5 percent probability lies with a minimum value of 10 million. Then you say that assuming this was a daily distribution then you can say on any given day your loss can be more than 10 million has a probability of 5 percent. Either it can be presented in terms of absolute value or also in terms of return form that is x percentage talking about negative returns or losses. So there are two ways to express this phenomena. One you can either say that there is a 95 percent probability that your maximum loss is 10 million dollars or there is extreme 5 percent possibility that your loss can be more than 10 million dollars.

So these are two ways to express the same thing and in that case it is said that your 5 percent daily war is 10 million dollar. There can be a parallel negative return in the form like x percent or losses that can also be expressed in war form. So let us say this 10 million war was x percent then you would have said that your 5 percent daily war is x percent which is negative losses or negative returns. Before I elaborate this formula in detail, please note that there are three important inputs to specify war. First, there is a time period.

In the previous case we noted daily period. It may be weekly or monthly as well. Second the level of confidence that is 95 percent or 99 percent in that case if confidence is 95 percent then that alpha value in the previous case it was 5 percent and if you take a confidence level of 99 then by the same logic your tail probability that you are examining will be 1 percent that is 1 minus alpha will be 99 percent then alpha will be 1 percent. And lastly the estimate of loss in absolute terms, this estimate of loss which was 10 million dollar in absolute term or in percentages term. Now from our knowledge of standard normal distribution tables we already know that if this point is supposed to be 1 point, is supposed to be 5 percent tail, the 5 percent probability on this line is 1.

645 standard deviation on the left. We can identify this point as 1.645 z value. So z value equal to 1.645 on the left side will correspond to that 95 percent region that means area on the left side is 5 percent and on the right side is 95 percent.

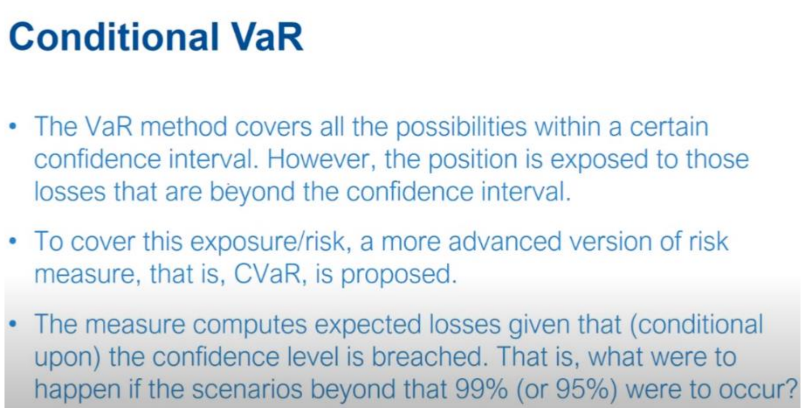
Now let us come to this formula. What does this formula mean? This formula is quite simple. It indicates the VAR value for a given distribution of returns. Let us say you want to have this alpha value or value of significance as 5 percent then the VAR value is defined as the VAR value is defined as if we plot the return distribution for that given frequency which is the area weekly the 5 percentile or 5 percent quantile or you can say 5 percentile region the minimum value in this region is the VAR value. Now if you have plotted the probability density of return distribution let us say and you have identified this 5 percent probability region the minimum percentile value will of course be that particular quantile or percentile.

For example we are talking about 5 percentile that means the minimum value will correspond to exactly this 5 percentile cut off point. This is what is represented by this formula that the minimum value for a given chosen value of alpha percentile or quantile let us say you are talking about 5 percentile then the cut off value is the minimum value of z that corresponds to this cut off. So which is nothing but simply the percentile value corresponding to the 5 percentile or alpha level. Now there are two ways to estimate this number. First is if you know the probability distribution for example if you believe that your variable or return follows normal distribution and you can plot them around in the form of normal or standard normal distribution then in that case you can simply integrate over this region.

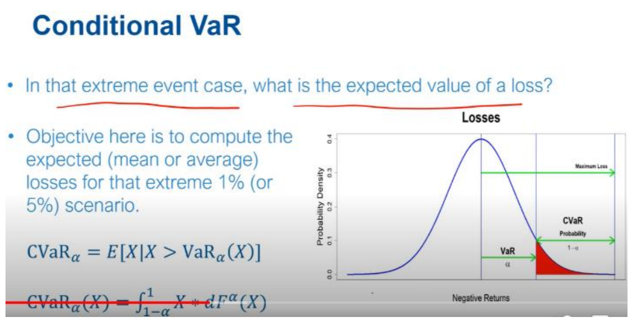
For example if I want 10 percent VAR daily VAR then I can find a cut off point on the left of which the area because we are talking about probability density curve then we can say the area under the curve is 10 percent of the overall area and in that case that integrated number because we want area under the curve that integrated part will represent the area the cumulative area the corresponding cut off point z corresponding to this 10 percent area will become my VAR value. So I can identify this z value here and this z is nothing but returns in standardized form so corresponding to each z value there is a particular return value which would be negative because this is the loss region on the left of 0 so we can identify that loss return and that will be our daily VAR with that particular 90 percent confidence or 10 percent significance or we can say 10 percent VAR daily VAR is this value. It can be daily, weekly, monthly depending upon the distribution of returns that is being examined. The second and empirical way is to have return data distribution data we can order this and start examining from the negative side negative return or the lower side and let us say if you are looking for 5 percent VAR then you can cut off the extreme negative 5 extreme lowest 5 percent observations you can cut off let us say you have set of observations you can put a cut off line and on the left of this cut off line you have 5 percent observations then the cut off point the maximum return in this cut off point becomes your VAR. In terms of magnitude this will be the lowest value on the negative side but in terms of signed value it will be the maximum and this will become our VAR.

For example let us say your values on the negative side are minus 5 percent, minus 4 percent, minus 3 percent, minus 2 percent and minus 1 percent. Now please note in terms of magnitude this 1 percent is the lowest however in terms of signed value this is the highest and this value once we cut off at this level and we have these 5 values left then this minus 1 percent which is the highest or in terms of magnitude the lowest value will become my VAR value. So we need to order our return observations and take the one which is highest in terms of magnitude or lowest in terms of magnitude and highest in terms of signed value. Mostly chances are it will be negative. So generally returns are symmetrically to some extent they are symmetrically distributed so you have negative as well as positive returns.

So if you are ordering them from the lowest side chances are that on extreme 5 percent or 10 percent you will have negative returns like this. This value at risk measure is often very often employed by banks to estimate the performance of their portfolios and possibility of extreme loss events. To summarize in this video we discussed the concept of value at risk and how it is helpful in estimating the tail risk of a security or a portfolio. However we noted that there can be instances that can and obviously will go beyond and extreme loss events that are beyond this VAR level and therefore we need a more slightly more accurate and efficient measure to handle those situations that are extreme situations and goes beyond this value at risk measure. In the next video we will discuss that one such measure which is called conditional VAR or C-VAR or expected loss.

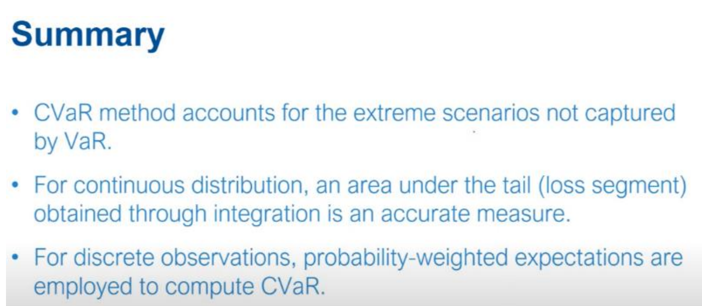


In this video we will introduce a very important measure of tail risk that is conditional VAR, C-VAR or expected shortfall measure. VAR method, value at risk method that we discussed in the previous video considers all the possibilities with a given certain confidence interval. For example we said we are 95% confident or 99% confident. However there are scenarios when this position is exposed to those losses that are beyond these levels of 95 or 99% confidence. And in order to cover this exposure that may happen in those extreme events a more advanced version of risk measure that is conditional VAR or C-VAR often referred to as expected shortfall is proposed.



$$CVaR_a = E[X|X > VaR_a(X)]$$

$$CVaR_a(X) = \int_{1-a}^{1} X * \, dF^a(X)$$



The measure computes those losses, expected losses given that this confidence level of 95% is reached. For example given that you are in this extreme region, given that you are in this region what is your expectation of losses? Let us say this is 99% or if you believe 95% then this region if it is 95% cut off then what is your expected loss if you are in this region or this confidence level is breached? In order to answer this question that we are talking about extreme event case what is the expected value of loss? To answer this question what we do here is we plot our returns on the positive side in contrast to our earlier plot where we had negative returns here and positive here. We essentially plot the losses that means negative returns so this will be our positive side and these extreme scenarios will represent those extreme negative returns or losses. Now objective here is to compute the expected or average or mean value of losses given that this these events were to happen or occur.

Let us take a simple case. Let us say you have three possible scenario given that this scenario is breached that is conditional sort of base probability that this extreme limit alpha it can be 5% that means 95% confidence level or 99% confidence level is alpha is 1%. This level is breached and we are talking about this region condition on that that a priori probability is given that this region is breached you have three possible scenarios P1 probability let us say 10% that loss is 5 million. Let us say I am putting negative sign for loss and another 10% probability that loss may be 10 million and 80% probability that loss may be 1 million. Then my expected loss given that these are the only three possibilities given that my losses exceed this level then P1 into this minus 5 P2 that is 10% into 10 and 80% into 1 will be my resulting number.

This is my 0.8 then this is 10 million so 1 and then 10% with 5 so 0.5 so overall my losses are 2.

1 plus 1 0.5 plus 0.8 that is 2.3 million losses. So these are my 2.3 million losses given that the losses exceed this level. So this is why this is conditional VAR because condition that this VAR level is breached so condition to that what are my expectation of losses. Now this was a rather simple case for discrete events and therefore it can be represented very simply like this that condition this is the conditional operator condition that X is greater than VAR that means my VAR level is breached that is my given condition what is the expected value of X and which we simply computed as P1 X1 plus P2 X2 and so on.

If your returns are in continuous form or you believe that returns follow some kind of continuous distribution then also the case is not too difficult you have continuous value of return observations you can plot them and for each return observation there is a small probability density value that this return observation will materialize or if you are talking about in terms of loss then X value along with their probability density and then you can integrate this is 1 minus alpha region please remember this is 1 minus alpha region which is your critical region your confidence level let us say here if you are talking about confidence let us say alpha we are denoting slightly in a different way let us say my alpha is my confidence of 95 percent so if this region is 95 percent this remaining tail region of 5 percent which is 1 minus 0.95 this 5 percent region I will integrate from this 1 minus alpha point to 1 so I will integrate from 1 minus alpha to 1 I will integrate this X value and df is my probability density which will also a function of X so I will integrate X df where f is function of X starting from this 1 minus alpha point assuming that this alpha is my confidence interval so from this 1 minus alpha to 1 if I integrate I will essentially get this probability which is my expected value of loss given that the value has exceeded this var level and it will be something like 1 minus alpha to 1 integrate X df X. To summarize this video we noted that C-var method accounts for extreme scenarios that are not captured by var levels for continuous distributions the area under the tail segment as we noted which is essentially the loss segment which was obtained through integration is a very accurate measure of that expected shortfall or conditional C-var measure for discrete observations also we noted that p1 x1 plus p2 x2 kind of probability weighted expectations can be employed to compute C-var given that the var levels has exceeded given that var has exceeded the a priori probabilities of these extreme losses p1 p2 p3 and so on can be multiplied by these extreme loss events x1 x2 x3 and these probability weighted expectations are used to compute our conditional var. As we start our journey of portfolio construction and its implementation with our programming in this video we will begin with downloading data from a very powerful and publicly available data source that is Yahoo! finance. We will start with downloading the securities first we will download the relevant packages starting from library command fportfolio that is the most important package for this then we have library tidyverse and we will keep these packages loaded for times to come so that whenever we are running a function related to portfolio construction or portfolio optimization we are able to run those.

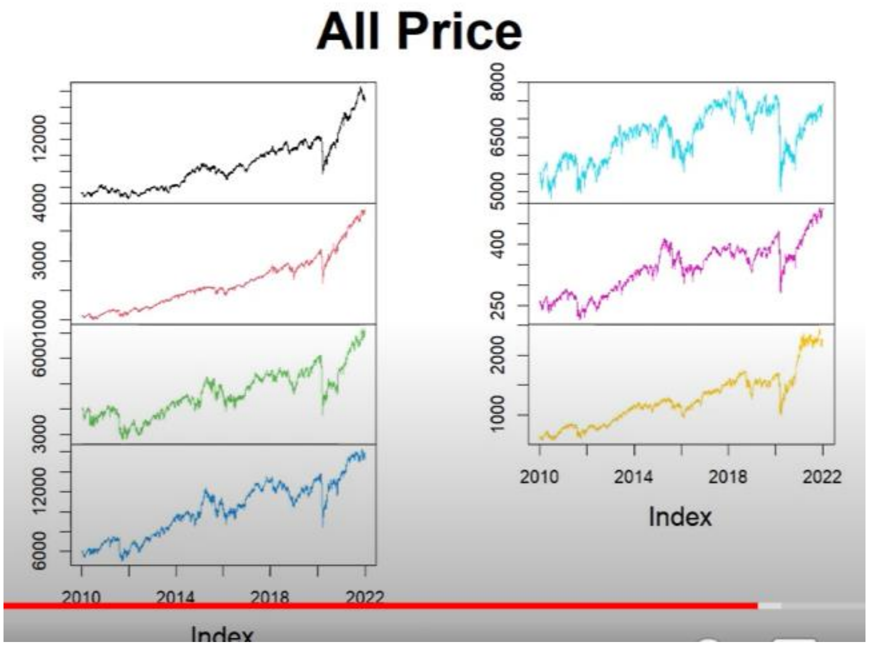
Another important package is library zoo we have library quant mode. Now that we have loaded the relevant packages we will start with the downloading securities. As a sample let me show you if I download a single security I need to know and please remember I need to know the ticker on Yahoo! finance in order to download the security I already know that Amazon has a ticker amzn so I will use this get symbols command

and this symbols I need to provide the symbol that I want to download which is Amazon. Now here I am downloading a single security I am downloading all the data in this amzn object and I will put this command auto dot assign equal to f and I will supply the period also from maybe 2010 first of Jan so this is my starting period and I will provide this data to 2021 December 31st so this is my closing period. Now when I run this command notice when I run this command amzn object is created on my environment so when I run this command notice as this command is run data from Yahoo! finance is loaded let me see let us examine this object so we have the following data which is open high low close volume and adjusted price is loaded so we will use this adjusted price series which include any adjustment for dividend and other cash flows so we will use this amazon.

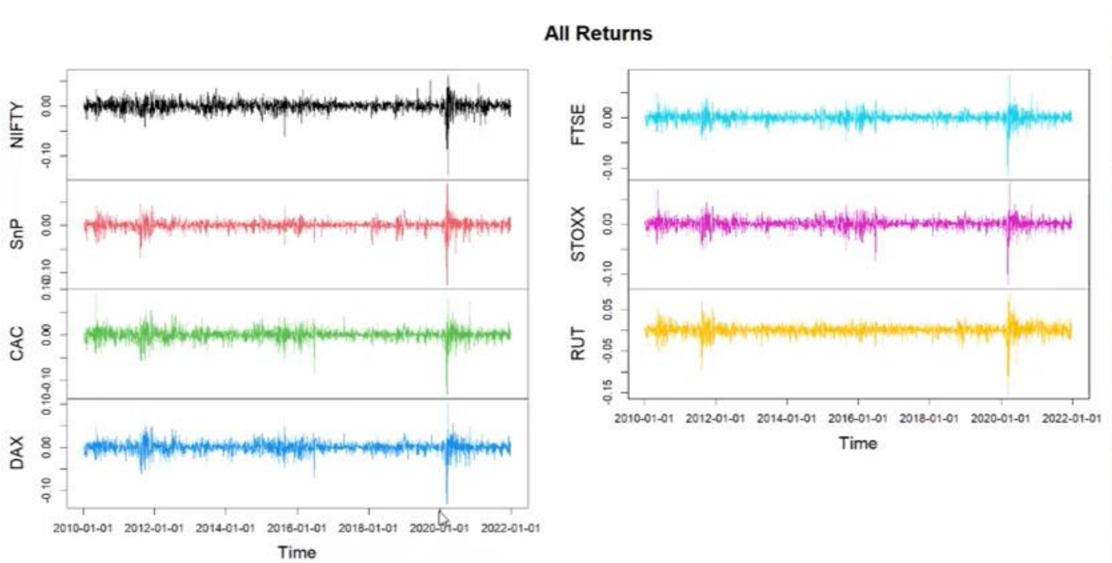
adjusted pricing. Let us examine the class of this object so this is XTS or zoo kind of object that is downloaded. Now please remember in order to construct portfolio we need to download multiple objects so for that again similar code is used let us call this multiple series object as df or data frame I will use the same command with same symbols get symbol this get symbols command I will use but this time around I will provide a number of symbols because we need multiple stocks to construct our portfolio so first this symbol this is for Nifty 50 and then a CI this is for Nifty 50 and it I will provide it in vector form. Next we have this JSPC this is for S&P 500 later I will specify their more accurate names to avoid any confusion. Then we have FSCHI which is for CSE 40 major index in France it is a major national index then you have GDEXI this is for DAX Germany major market index in Germany.

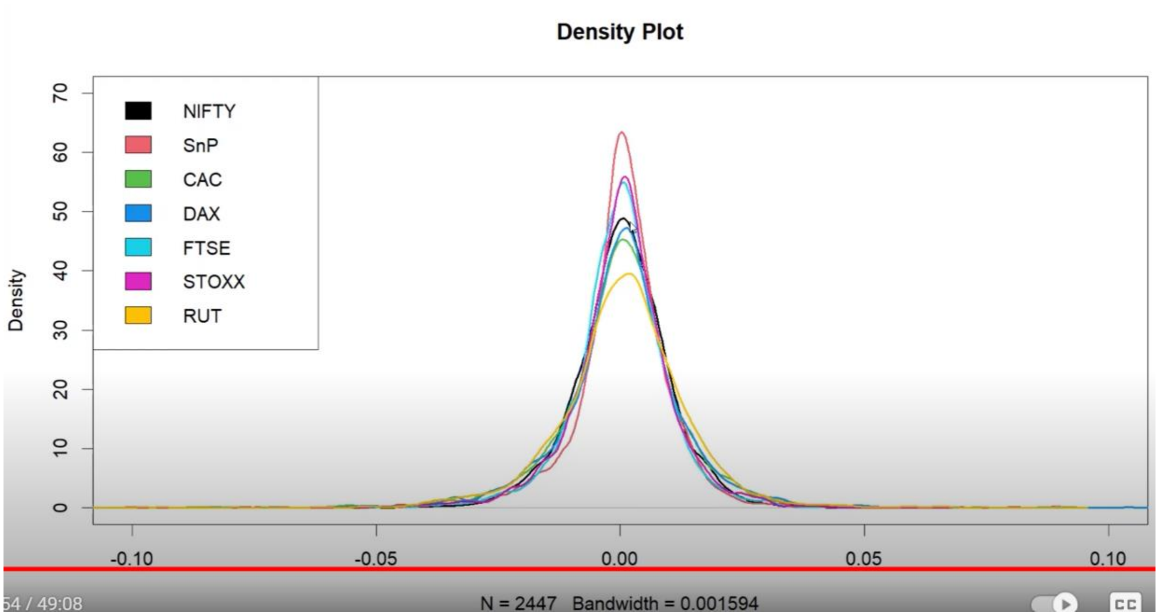
Then you have FTSE this is for FTSE 100 in UK. Next you have stocks which is major Euro stocks index and then you have Russell which is another Russell 2000 which is another major index for this which we have RUT so these are the major indices that we have signed. Now interestingly in this particular command where I am loading multiple securities I will not use auto.assign as false I will set auto.assign as true so all the symbols will be separately loaded in separate objects not assigned to this TF and I will use the same time window the time window I will keep as same which I used earlier and now I will load this. Notice when I run this command notice when I run this command in my data environment all the objects are loaded separately and they contain each object contains the data and now I need to combine the entire data object so I need to use the tickers by the name they are loaded ticker let's call them tickers and I will use these tickers as list objects combine them first is NSCI then I have JSPC these names are the names by which they are loaded in my environment I will change them to more suitable names shortly GDEXE FTSE these are my objects I am using major market in D6 these are generally less noisy so portfolio construction is expected to be smooth without any hazards of thin trading issues related to thin trading and so on.

So now I will run this command and load and create a tick this object tickers which is a list object I will merge and this ticker if I print this ticker object notice it carries the consolidated thing from this consolidated thing I will merge these are all in the list form so each object is contained individually so I will bind them together let's create a new data frame DF1 which will essentially bind all the objects let's use this merge command to bind them it will bind all the objects once I have binded in DF1 notice a large extensible time series XDS object I will select only adjusted prices and how do I select only adjusted prices it's quite simple I will use this final under let's call it final underscore DF where I will filter the DF underscore 1 series which is my combined consolidated series I will use this string detect command and in this string detect function I will supply the column names of this DF1 object and I compare it with the thumb adjusted column those columns or variables where this adjusted term is there they will be selected and the meaning will be ignored and if I have done it correctly let's let me look at final dot underscore DF object and let's see what kind of columns are there notice all the indices only their adjusted prices the column with the adjusted prices are appropriately selected so this is my final data to summarize this video we loaded the relevant packages of F portfolio tidyverse zoo and quant mode then we demonstrated how to download data for yao finance using a single security that is amazon after that we downloaded a number of securities from yao finance that is nifty 50 s&p 500 csc 40 from paris dax germany 4 C 100 from uk stocks from euro and result 2000 and then we combined them extracted their adjusted prices in our consolidated data frame format which is final underscore DF and this will be further used in the subsequent processing and return computation in the next video we will compute the returns and we will try to visualize the data now that we have downloaded the data let us start by properly naming it naming the variables or the security prices and then computing returns and subsequently visualizing it so as a first step we will give appropriate names to the column so we will name compute returns and visualize so the first step let me just give the names with this call names command here i'll use final underscore d this was our original data that we downloaded but we give it appropriate names the first name is nifty so since we downloaded nifty 50 data i'm giving it a name nifty then s&p 500 let me put it lsp or more appropriately smp next column is csc then dax then ftsc next we have talks euro stocks first box then result 2000 let me call it rut that is the cell 2000 so this is my final variable let me name these and if you want to check whether names have been appropriately given you can again type final in this book df and you'll find that if i enlarge my console window you can notice the names that are given are here visible as a next step let me compute the returns and a very simple interesting command let me call the new data as final underscore df underscore right so this command will compute it returns and it's very simple diff log this is a very simple command it takes the first difference of the series in the log form which is essentially logged returns this is very easy to do those of us who have computed return in the past so we'll those of us who have computed returns in the past they will know that this command will first create the log of the price series and then take the difference which is sort of returns in the continuous form and now that we have this return variable available with us let's see the head of this final underscore diff object that we have created let's see the head of this object yes the returns are computed the first object is obviously na because we use the first price series there is no previous price so the first prices are in it we can in fact to avoid any kind of trouble we can just type this command and we'll use this na.omit to remove na's from this object so I'll remove the na's from this object now na's are removed also this f portfolio object or library that we installed it generally works well with the time series objects there is a certain time series format that it's most suitable so we'll run this time series command and this time series command will convert our final diff rate object to essentially to this final return object let me name it as final rate in fact so this is my final return object that I'll be using so now that we have downloaded and combined our data into final return form let us start with a little bit visualization of the data so we'll visualize the data now to summarize in this video we gave the names to variables so we appropriately named the variables with their proper security names and then using diff lock command we computed the returns and then we removed any observations and then we converted our return object to a more suitable time series object in the next video we'll try to visualize data and see its properties in this video we'll try to visualize the data so we'll try to visualize the data let us start with a simple visualization command which is plot dot XTS and we'll supply this function plot dot XTS our final data frame which is final underscore DF this data contains prices we'll supply since we have seven price series seven indices we'll use seven colors so this command will ensure that there are seven colors one to seven then central heading is all prices because we are using price series and then we'll keep the while lab five X's as empty so now I can run this command notice notice that once I run this command I get all the price series on my chart I can see all the prices also you may want to plot all the series individually rather than putting them in single plot so that you can make a more appropriate comparison and see their variation properly so for that you can slightly more tweak the command for example you can type plot dot zoo and then put supply the same object final score DF you can again supply all the different colors since we want to plot all the prices in different colors again we'll put the central heading as all price but this time we'll specify that plot dot type is to be plotted in multiple windows so we'll specify this multiple argument so that plotting is done with multiple windows also we'll keep the y-axis as empty and then we can use this six dot main argument to increase the size of central heading probably a larger central heading will plot notice now all the prices are plotted with different different colors and different plot windows you can see their movement more clearly than on the single plot they are always well you can zoom it also you can create so the prices and see sometimes you may be not very sure that which window plots which which color or which window plots which series so you can remove this while app which was empty and now if I run this command now this command is run where I have not specified the y-axis so the original name from the series will be picked up let me just enlarge the plotting window so that we can see the graph appropriately so if I run this command notice the northern name of the axis that I specified earlier is also provided so for example nifty is the black one snp in the slightly red color reddish color cic is green and so on for example Russell is in sort of brown yellowish color so we have all the indices that we can identify individually in fact on all the seven indices individually on the graphs let's try to plot returns now we use make use of same commands it's quite easy so now it would be more appropriate to plot the return diagrams written graphs on separate windows because return is very fluctuating around mean of zero it generally it is very fluctuating so it would be appropriate if we plot them on separate plots so again I'll use the plot dot zoo but I'll supply the final return argument here we have the returns all the returns available again all the seven will be plotted in different colors central heading is all returns because this graph carries returns we'll specify the plot plot type plot dot type as multiple so that they are plotted in multiple windows and I'll not specify the y-axis so that each lot has its own index or return name I'll just specify that six dot main equal to three so that slightly larger graph is produced so now let me enlarge the plot window before I run the command so sufficiently I've enlarged it now once I run it notice all the individual indices are plotted and we can see the fluctuations in fact we can see around 2000 the around somewhere around after 2000 there's a large jump you can see here this large jump and this is common to all and in between also the fluctuations happen because these are broad market by indices so whatever whenever there's some kind of global crisis there would be some kind of co-movement high volatility and falling prices which is what we are seeing here and all the indices are plotted in different colors also a slight modification the command if I remove this rule notice if I remove the zoo now the dates are plotted so I can now I have all the individual indices and their dates are also available with me so I can identify these fluctuations their timings you can see there these timings are around somewhere on 2020 there is a sharp increase in volatility particularly the down which is also associated with fallen negative falling negative returns so we can say this is probably around covert period and it is common to all the indices as a last step when you are summarizing the data you've already plotted it you have visualized it as a last step it is always advised to create a summary measure so just summarize your data by tapping summary and supplying your final return object so it's nice that it will give you the final summary thing you can see that summary of all the returns is provided here nicely printed somebody you have start and end date of the record and you have nifty S&P the summary returns are also provided nicely printed somebody maybe can see that extreme negative returns are always mostly dominating the returns mostly these series are negatively skewed as it appears to summarize this video we plotted the security prices together in a single graph then we also plotted them in separate separate graphs then we plotted the returns we can



see that returns are negatively skewed especially during the covert time there was large fluctuation and negative returns this we confirm with the summary measure we summarize the returns and saw their distribution minimum maximum mean median first and third quantiles we observed extremely negative returns are slightly dominating so the return appear to be skewed we'll conclude with the visualization in the next video where we'll examine the density distribution of all the returns in the previous video we visualize the price and returns plot in this video we'll create density plot density plots to examine the distribution of the data as a first step we create the main plot of density so first we'll create the main plot of nasty this we'll do for nifty so nifty will be our first stop where we create the nasty plot we give the main heading is density plot also specify the color as one I'll give the color in sequence of one to seven there are seven securities next based on my experience I know the nature of plots while specify the X limits from minus 0.1 to 0.1 so they remain really aesthetic and Y limit as 0 to around 70 so the some of the big security then speed are also fairly plotted and the line width of two so this will be my first plot let's see if it is well it seems to have worked in addition to this I will create other plots but I will not create fresh but I'll use lines command to create new plots in the same plotting window let's see how to do that so I'll remove my main plot heading so this is not required I just need to change the name of the security this time it will be s&p 500 I'll change the color to two color coding scheme as two and I need not specify its limit in by limits that I have already done I only need to specify the line with has to nowfor next lines I just need to change the name of the security let's say s&p to CAC color as three for next security again from CAC to DAX color as four from DAX to FTSE color as five from FTSE I will shift to stocks color as six so I'm just changing the color scheme and lastly from stocks to Russell 2000 and color as seven so now these are my line plots for which I just add the main diagram notice all those plots are plotted in fact the red one is extremely big now you also need to add some kind of legend to make distinction between different plots so it seems that top left window slightly open so I'll use top left window to add the legend and I'll use a small trick and not explicitly enter the name but I'll use this call names command and extract the name from the return object itself lastly since I'm using the color coding from one to seven I'll continue to the same and then I'll add the legend so notice nicely legends are added probably I'll enlarge the plot window and then recreate this plot for better comprehension so now it looks better let's re-examine the plot so I'll zoom it notice all the plots in different colors for nifty in black s&p is more reddish CAC is green and they are plotted together in fact the red one seems s&p slightly more peak nifty is not so peak while Russell is much lower in terms of peak however if you look at overall the overall nature of these graphs it seems that broadly they can there can be a normal curve that can be superimposed on these they appear to be fairly bell-shaped and normal distribution and that is the reason earlier when if you recall our discussion on normal distribution with the topic we discuss in great detail how normal distribution is extremely useful for modeling such return behavior and we can see that here manifesting itself we can see how normal distribution nicely fits or appears very similar to these security distribution density plots to summarize this video we plotted the density distributions of all the seven security prices and we find them very similar to normal distribution you





Density Plot
NIFTY
SnP
CAC
DAX
FTSE
STOXX
RUT
Density
0
10
20
30
40
50
60
70
-0.10
-0.05
0.00
0.05
0.10
N = 2447 Bandwidth = 0.001594