Advanced Algorithmic Trading and Portfolio Management

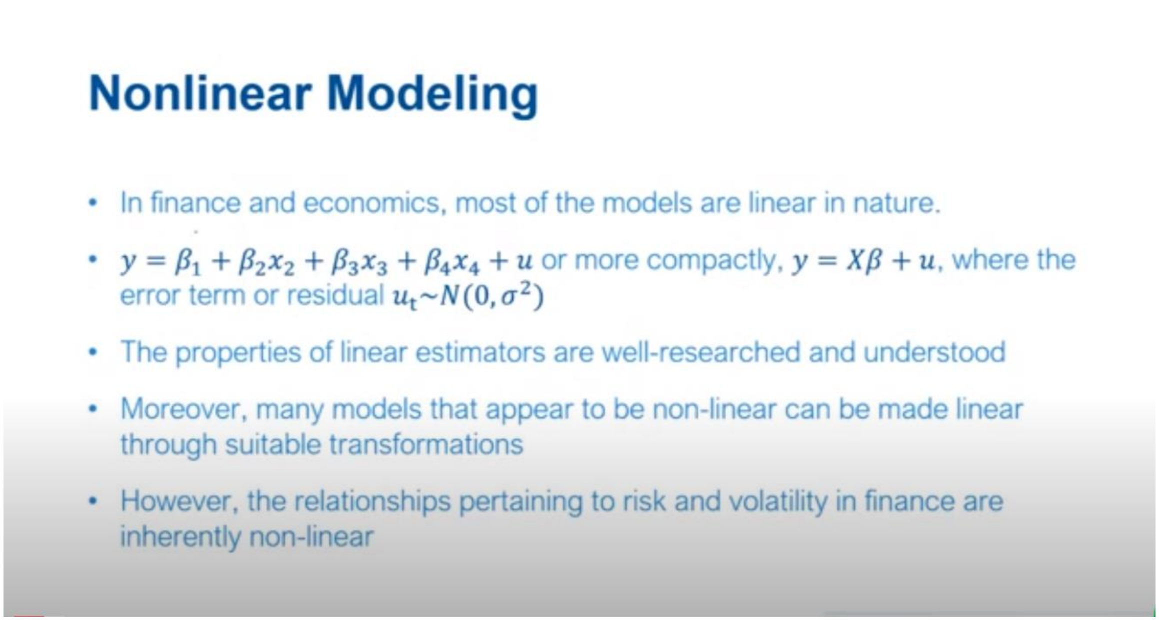
Prof. Abhinava Tripathi
Department of Management Sciences
Indian Institute of Technology, Kanpur

Lecture- 27

In this lesson, we will introduce various volatility models such as exponential weighted moving average models, autoregressive conditional heteroskedasticity ARCH models and generalized ARCH models. We will start the discussion by providing some stylized empirical phenomena that result in nonlinearity in the relationships pertaining to volatility and risk and therefore the subsequent requirement of nonlinear models in the context of risk and volatility. These phenomena include negative skewness, excess kurtosis, volatility clustering and leverage effects often associated with volatility in financial markets. We start by introducing historical volatility models and implied volatility models. Next we discuss conditional volatility models. Then we discuss the shortcomings of these models that result in more advanced models such as EWMA or exponential moving average model.

The phenomena of volatility clustering leads to ARCH class of models. However, the ARCH class of models are less parsimonious. This requirement subsequently leads to more sophisticated GARCH family models. In the GARCH family model, we discuss standard GARCH11, EGARCH and GJRGARCH models.

In particular, these advanced GARCH models are extremely useful in capturing the volatility clustering and leverage effects observed in financial markets. In this video, we will study the background and motivation behind studying nonlinear models in the context of volatility and risk modeling.



In finance and economics, most of the models are linear in nature. For example, have a look at this model

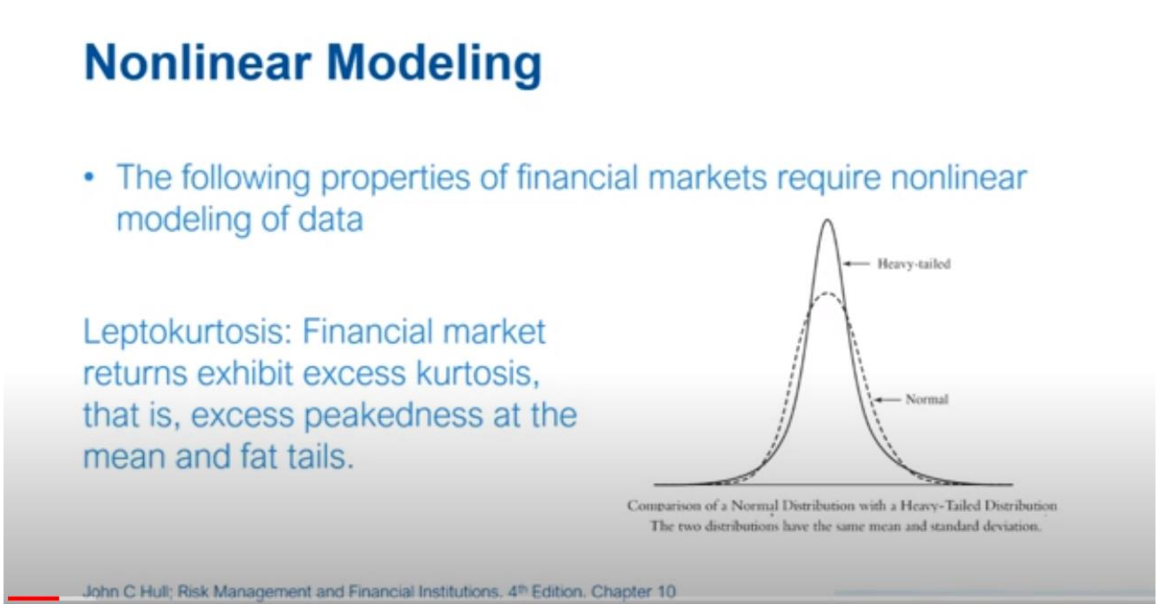
$$Y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \mu$$

and then this error term mu are more compactly in a matrix form

$$y = x\beta + \mu$$

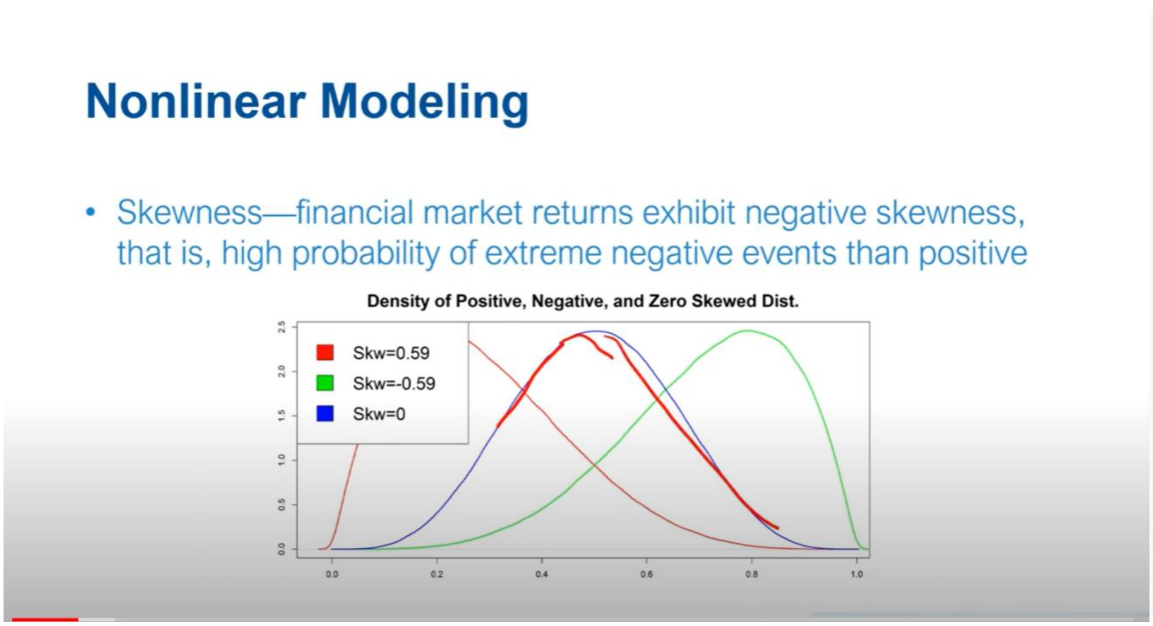
where error term is often assumed to be distributed with a normal distribution and zero mean and variance of sigma square. The properties of linear estimators such as OLS to estimate this kind of linear model are well researched and understood.

Moreover, many models that appear to be nonlinear in nature can also be made linear through suitable transformations. For example, a model like Y equal to alpha plus beta X square, we can take the log of for example Y equal to alpha plus alpha into e to the power beta X. Such model can be made linear by taking log transformations. So there are number of ways through suitable transformations such relationships can be made linear. However, in particular the relationships pertaining to risk and volatility in finance are often considered for nonlinear modeling.



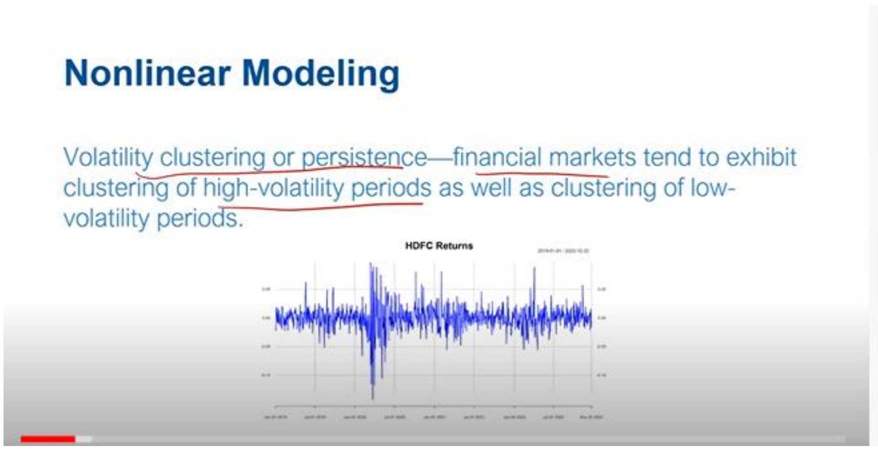
Let's explore why. One of the very important property of financial market returns is called kurtosis. While the normal distribution assumes a kurtosis of 3, the financial return data often does not agree to that. For example, it often exhibits something called leptokurtosis where it exhibits excess kurtosis that is excess peaked Ness for example as compared to the normal distribution here in dotted form. The actual return distribution may be excessively p and also exhibit fat tails.

So it exhibits fact tails that means higher probability in extremities, extreme probability as compared to that predicted by a normal distribution. So, it has evident fat tails and excess peakedness which makes it difficult to model through linear relationships that assume normality.



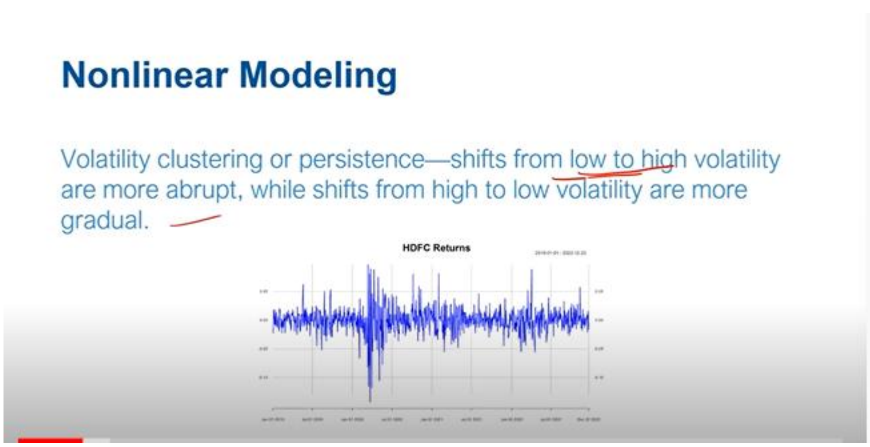
The next important property studied in financial markets is the symmetry of distribution. And often it is found that unlike the predictions of normal distribution which says for example here in the blue curve, it's very symmetric on both sides, positive and negative. However, actual financial market returns exhibit negative skewness.

That means there is a high probability of negative events, negative events as compared to positive side. So, this is skewed sort of skewed distribution and it is often negatively skewed. So, there is a high probability of negative returns being observed.

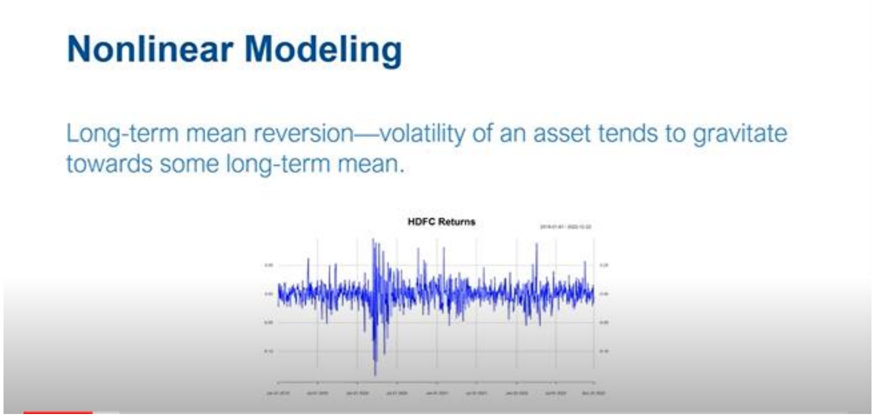


Another very important property is volatility clustering or volatility persistence. That is in financial markets, you tend to find clustering of high volatility periods as well as clustering of low volatility periods.

For example, here you can see clustering and munching of high fluctuations and clustering and munching of low fluctuations together which suggests that period of high volatility occur together while periods of low volatility occur together. This kind of phenomena is slightly difficult to model through linear relationships as it indicates some kind of autoregressive nature in volatility.

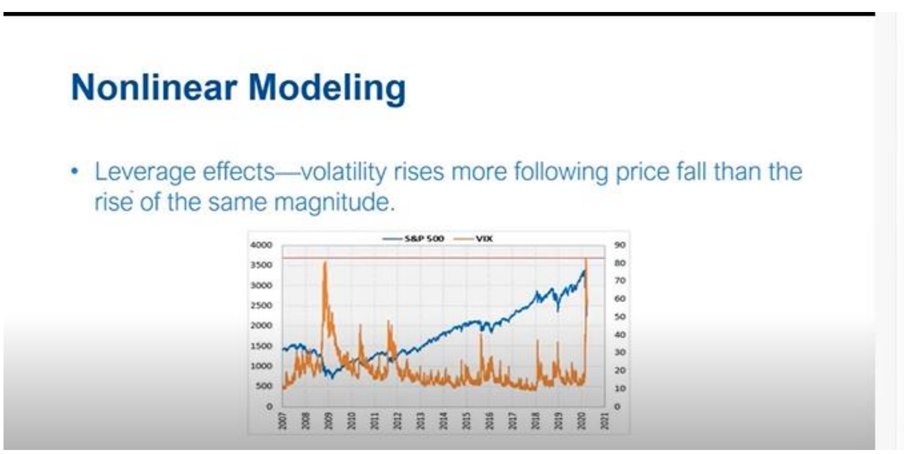


Another stylized fact of volatility clustering and persistence is that shifts from high to low and low to high volatility are not uniform. That is low to high volatility shifts are more abrupt. You suddenly find jumping from low to high volatility while shifts from high to low that is in this fashion are more gradual. So, shifts from high to low volatility are more gradual while low to high volatility are more sharp.



The next property which is very critical is called long term mean reversion. That is volatility of assets tends to gravitate towards long term mean. Generally this long term volatility level is sort of inherent level of volatility which irrespective of whether there is some information or not or whatever market conditions, this long term mean volatility always persists. Researchers believe that this is driven by microstructure volatility.

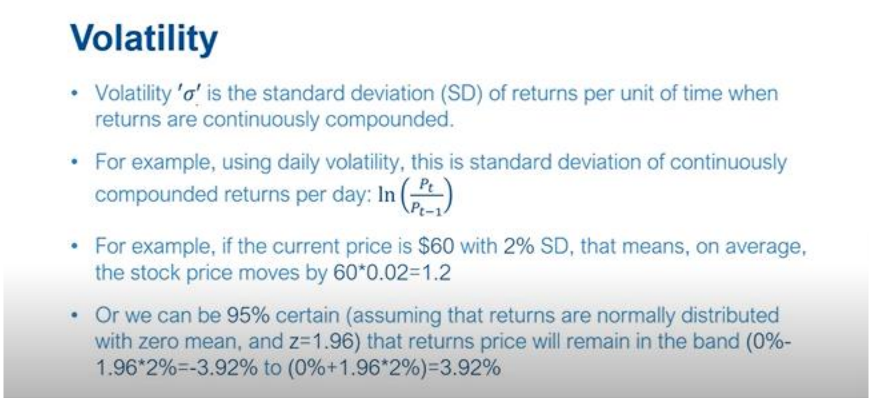
For example bid-ask bounds. So, phenomena that are part of market microstructure such as bid-ask bounds lead to this all-time inherent volatility of financial markets.



Lastly a very stylized fact is leveraging effects. It is often observed that volatility rises more following periods of falling prices and as compared to that when prices are rising volatility is less.

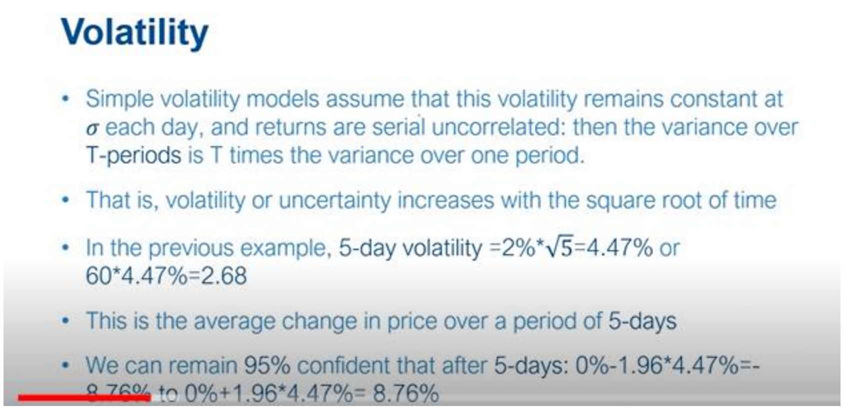
For example, during the rising periods volatility is less and falling periods there is a higher volatility. It is often argued that when prices are falling the leverage that is debt to equity ratio increases. Inherently debt to equity ratio is a measurement of risk. A higher debt to equity ratio indicates a high-risk level and therefore there is a feedback sort of loop created. That means if prices are falling and debt to equity ratio is rising inherently company becomes more riskier and therefore this further contributes to additional risk or additional volatility to the already existing levels of volatility and therefore it is said that volatility rises more falling price fall than as compared to that during the rise of the price of the same magnitude. To summarize in this video, we discussed some of the stylized fact of volatility and risk that require us to model the relationships near to volatility risk for nonlinear modeling. For example, we said that volatility in financial markets or distribution of returns exhibits excess kurtosis as compared to normal distribution. It is negatively skewed. We also observed volatility clustering, that is high periods of high volatility are clustered together. We also observed what we call leverage effects, that is volatility rises more during falling prices as compared to that during rising prices. We also discussed that there is some kind of long term mean reversion property of volatility and also the transition to high volatility periods from low volatility periods are more abrupt while those from high volatility to low are more gradual.

In this video we will discuss volatility and its theoretical underpinnings behind the computation of volatility and as a mathematical measure to proxy risk.



Recall that volatility or denoted by sigma is the standard deviation of returns per unit of time when returns are continuously compounded. For example while we are considering daily volatility this is the standard deviation of continuously compounded returns per day. So the formula for continuously compounded returns is natural ln $(\frac{Pt}{P_{t-1}})$ and when you take the standard deviation of continuously compounded returns for a period let us say our day then that number is the standard deviation of daily returns. Let us take an example.

For example if your current price is sixty dollar and it is given to you that standard deviation or daily volatility or standard deviation daily standard deviation is two percent that means on average the stock moves by up and down by one point two dollars on a given day, so average movement. Or in a more concrete manner if you want to translate your understanding through a probability distribution like normal distribution let us say you believe that returns are normally distributed with a mean zero or zero percent and you want to know the ninety five percent confidence interval for return these two cut offs if you believe that returns are normally distributed then these ends ninety five percent ends these markers are one point nine six minus one point nine six and plus one point nine six that is two point five percent probability here extreme probability two point five percent probability here extreme probability and in between you have ninety five percent confidence interval. If that were to be the case then your window of ninety five percent interval of returns is one minus one point nine six zero percent minus one point nine six into two percent this is on the lower side of it which is minus three point nine two percent the upper side is very also symmetric so it is zero percent plus one point nine six into two percent which is again three point nine two percent. So effectively if you want to know that the range of the price that will be sixty dollars into one minus three point nine two percent and sixty dollar into one plus three point nine two percent. So, this will be your price lower range lower level and upper level of prices with ninety five percent confidence band.

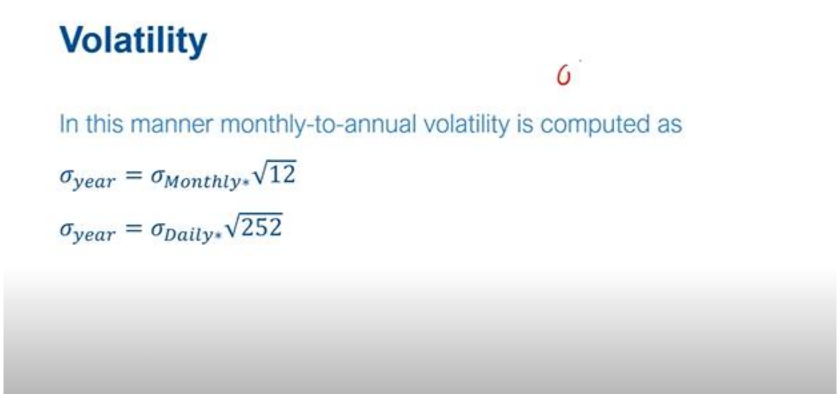


Another very important property of volatility that is driven by simple volatility models is that they assume that the volatility remains constant at each day. Basically the assumption is that your returns are independently and identically I distributed. So returns are IID that means they are independently and identically distributed which is to suggest that first the distribution remains same that means the standard deviation or volatility remains constant for each period each day at some level sigma. So whatever this level is it is constant each day and the distribution remains same and also returns are certainly not correlated. If these properties are held then a resulting property is that variance over t periods is t times the variance of one period or in other words volatility when measured through standard deviation it increases with the square root of time.

Let's understand what it means. In the previous example we said that daily volatility was two percent and we want to convert into five day volatility which means five periods. So what we'll do is we'll multiply the standard deviation value of two percent with the square root of the period t which is five. So we get a value of four point four seven percent or in dollar terms sixteen to four point four seven percent which is two point six eight. What it means now that if you want to know the average expected change in a period over a period of five days then it can be either plus minus two point six eight dollars it can go up or down by two point six eight dollars.

Let's also think in terms of what it means if you want to have a ninety five percent confidence interval. Again recall that we said these lower and upper ends of this if we believe that follows a normal distribution then these ends are one plus minus one point nine six and if the mean the returns are distributed with a mean of zero percent then the lower end of the return is zero percent minus one point nine six into four point four percent which is eight point six percent. Similarly the upper end is zero plus and with symmetry it is again plus eight point seven six percent. It was minus. So now if I want to know the average movement in my return the price return we saw now if I want to know the price it will be simply $60 * (1 - 8.76\%)$ and upper level would be$60 * (1 + 8.76\%)$

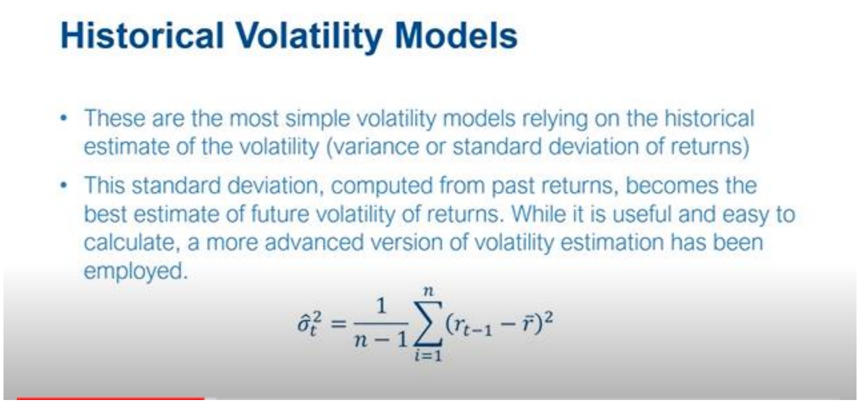
This would be my price bank for a five year period with an expected up and down moment and return of eight point six seven six percent with ninety five percent confidence level.



To put it more succinctly let's say if you have monthly standard deviation Sigma monthly is given at some level if you want to know the annual standard deviation you want to analyze it you simply multiply it by square root of twelve where t is equal to twelve periods. Similarly if Sigma daily is given to you you analyze it by multiplying number of trading days which is 252 trading days so you multiply it by square root of 252. To summarize this video we discussed the mathematical underpinnings behind the computation of standard deviation or volatility as a risk measure. We noted that it is the standard deviation of continuously compounded returns for a given period and to translate it to different periods we multiply the standard deviation measure by square root of the time periods.

For example if you wanted to convert the daily standard deviation to annual you multiply it by number of trading days in the year which is square root of 252 and so on for monthly and weekly periods. However this result is given by a simple assumption that returns are IID that is they are identically and independently distributed that is there is no serial correlation in returns across period by period and also the distribution of returns period by period remains the same which means the standard deviation is same at some level Sigma which does not change over period by period. However this seems to be a slightly difficult assumption to sustain.

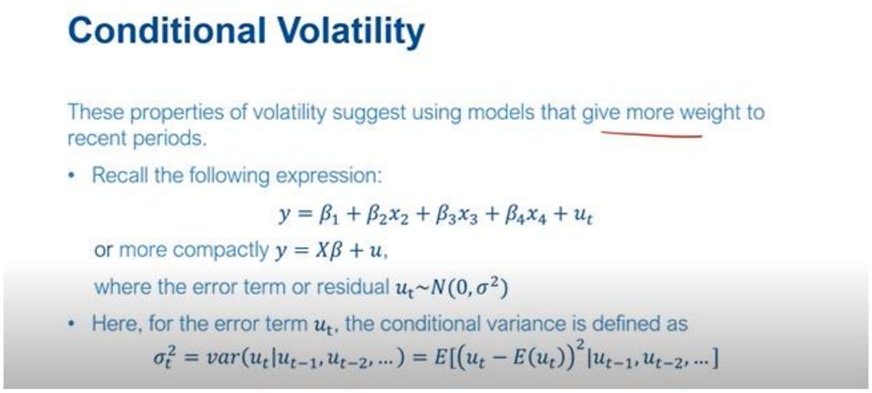
In this video we will discuss two very fundamental models of volatility that is historical volatility models and implied volatility models. While these models are very simplistic and obviously they face certain challenges in their assumptions still they provide the backbone and fundamental building blocks to more advanced models that will be discussed subsequently in this lesson.



To begin with we start with the historical volatility models. These are the most simple volatility models and they rely on historical estimate of the volatility or what we call as variance or standard deviation of returns. The standard deviation is computed simply from past returns and it becomes the best estimate for future volatility of returns if certain assumptions such as returns are independently and identically distributed or they are IID holds true. This model is very useful and easy to calculate for example the variance is nothing but deviation mean deviation squares divided by n minus 1. In some theoretical values you use n while n-1 is more of a sample characteristic to get a more unbiased estimate of volatility because generally it is assumed that we are working with samples not the population of returns.

However as you would have guessed it this model has a very fundamental problem that it gives equal weight to all the historical returns which is not a very good property. Generally you would like to give more weight to more recent observations of returns. The next set of models are implied volatility models that rely on some kind of option pricing formula such as Black-Scholes. So for example any option pricing model would use volatility as input along with some other parameters such as risk-free rate, strike price and current value of underlying to compute the option prices. The idea behind these implied volatility models is to use the observed value of parameters such as risk-free rate, option prices, strike prices and so on to back calculate the implied volatility from these models.

Now essentially this implied volatility is the market forecast of the volatility of underlying asset returns over the lifetime of the option. To summarize in this video we discussed two very fundamental models that is historical volatility model and implied volatility models. While these models are simple to understand and very intuitive the assumptions behind them are not very tenable in real life phenomena but still these models are widely held and employed because as such they provide the very fundamental building block to more advanced models as we will be discussing in subsequent videos. In this video we will introduce the concept of conditional volatility which is based upon the fact that more recent factors or more recent information arrival has more influence on the volatility levels observed currently. The idea behind conditional volatility models is to give more weight to recent periods.



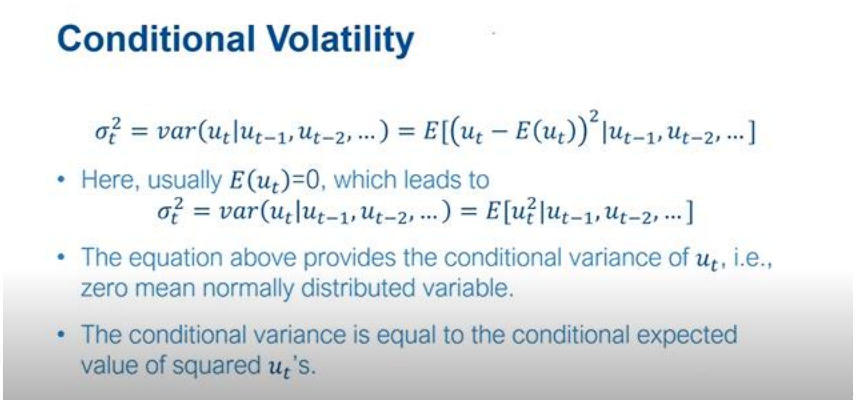
Let's start with the mathematical or theoretical underpinning behind this idea. Recall we said some kind of relationship in financial markets like

$$Y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \mu$$

Y equal to beta 1 plus beta X 2 and so on with an error term mu t is expressed like this where in vector or matrix form you can write $y = x\beta + \mu$. Generally it is assumed that mu t or error term or in financial markets the information term is normally distributed with a zero mean and constant variance sigma square. Now if you want to know the conditional variance of this error term it is simply variance of mu t given mu t minus 1 mu t minus 2 and so on that is conditional to the historical information like mu t minus 1 mu t minus 2 mu t minus 3 what is the variance of this mu t so that is conditional upon historical information rivals which is also written as expected value of mu t minus mu expected mu t square which is the nothing but mean deviation square what we have already seen earlier mean deviation squares of mu t conditional to previous historical information levels like mu t minus 1 mu t minus 2 and so on.

$$\sigma_t^2 = var(\mu_t | \mu_{t-1}, \mu_{t-2}, \dots.) = E\ [(\mu_t - E(\mu_t))^2 | \mu_{t-1}, \mu_{t-2}]$$

Now let's translate this understanding in the context of financial markets.



Imagine a relationship like this in financial market like return at time t is being regressed on historical returns lags like r t minus 1 plus r t minus 2 and so on up till certain period plus a certain error component. Please note in this kind of model the error term mu t would indicate the information that has arrived in the current period so after accounting for all the lags of returns or serial correlation returns whatever left in mu t is nothing but the latest information and therefore it is often called a measure of information or innovation which proxies the latest information arrival. Now the conditional variance of this mu t or sigma square t is what we call as conditional volatility or conditional variance which is conditional upon historical information or innovation such as mu t minus 1 mu t minus 2 and so on. As we have said earlier it is nothing but expected value of mean deviation squares that is mu t minus expected value of mu t whole square conditional upon mu t minus 1 mu t minus 2 and so on. Now recall assumption that this expected value of mu t is 0 so we can simply write the sigma square t as variance of mu t conditional upon mu t minus 1 mu t minus 2 and other historical information points which is nothing but expected value of mu t square or simply the variance of current period innovation or information that is μ_t^2 given the information or conditional to information that is μ_{t-1}, (μ_{t-2}), and so on.

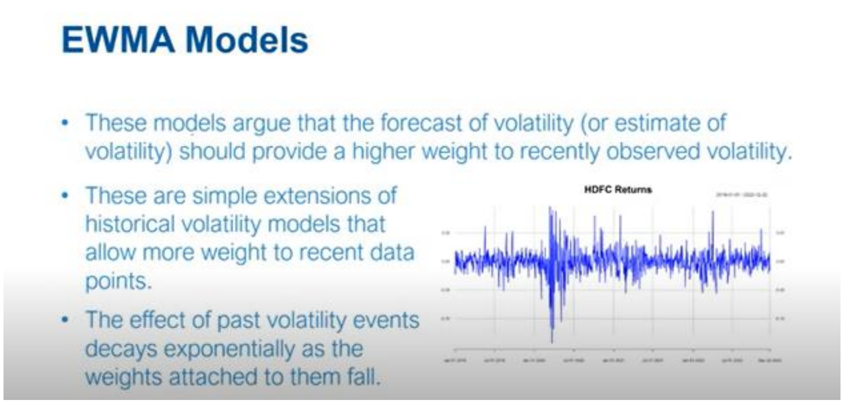
The above equation that we discussed provides the conditional variance of $\mu_t$that is zero mean normally distributed variable. The conditional variance is basically here equal to conditional expected value of squared mu t's. Now let me give you the intuition behind conditional volatility. Think of a string which is held at two ends A and B and it is very tight a jerk is given at end A so it starts fluctuating. The jerk was given at time t equal to t naught and they start fluctuating as time passes the jerk becomes smaller and smaller and smaller.

Suddenly at time t equal to t1 you give another t equal to t1 you give another jerk and again the fluctuation starts. Now please recall there was already some effect of the previous fluctuations which will still sustain and added over to it the jerk or hit that you gave at t equal to t1 that also contributes to it. So there are essentially the historical fluctuations in the string and more recent fluctuations. Obviously the impact of more recent fluctuations will be higher as historical fluctuations will die away. But if I want to know what is the impact of this latest hit or latest information or latest jerk to the string I need to somehow model out I need to somehow model out the historical jerks or historical fluctuations.

Once I model out or extract the historical information or fluctuation impacts on the string then only I would be able to measure the impact of this jerk or information that came information shock that came at time t equal to 1 and that is what we mean when we say that what is the variance conditional to the previous historical information shocks or information arrivals that is mu t minus 1 mu t minus 2 conditional to them what is the current variance or impact of shock that is mu t. So this is how we develop the intuition. To summarize in this video we saw the intuition and understanding of conditional volatility. The concept of conditional volatility derives from the fact that recent periods or recent information arrival has more say on the volatility levels and we need to in order to understand this we need to model out or sort of extract the historical information shocks so that we get a more pure and more sharp measure of recent volatility conditional to historical information shock or information arrivals.

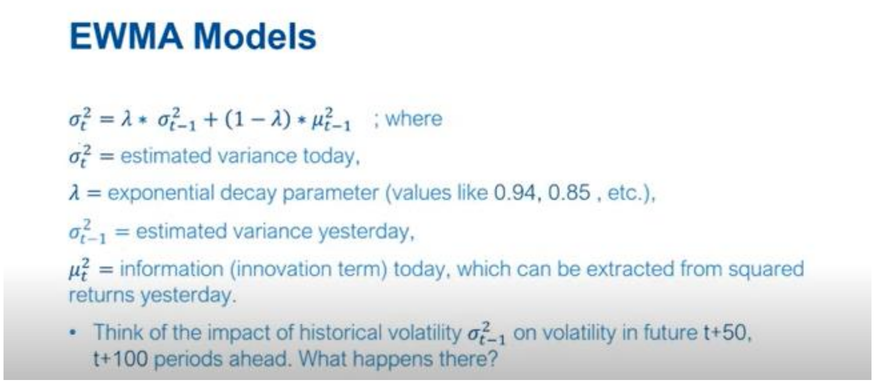
In the previous video we discussed historical volatility models and we said about these models that they give equal weight to all the historical observations no matter how fast or further in time this information may be.

Obviously this kind of mechanism is problematic and this leads us to what we call as exponentially weighted moving average models. In this video we will introduce EWMA or Exponentially Weighted Moving Average models and see its mathematical formulation and theoretical independence behind it.



To begin with these models argue that the forecast of volatility or estimate of volatility should provide a higher weight to recently observed volatility or observations while a lower weight to those observations that are much further in past. These are simple extensions of historical volatility models that allow more weight to recent data. The effect of past volatility event decays exponentially as the weights attached to them fall.

Now we will see how this works mathematically but just to give you some intuition suppose the given particular period end is in past t-n σ_{t-n}^2has a particularly high level of volatility. Now as per the historical volatility model as long as this model this particular observation is included in the model it will lead to heightened level of volatility. However even for a single period as soon as it goes away its exclusion would lead to sudden fall in volatility levels. Such drastic fluctuations in volatility forecasts are not desirable and therefore EWMA models precisely account for this fact by exponentially decaying the impact of this particular level of innovation or volatility on the volatility estimates.

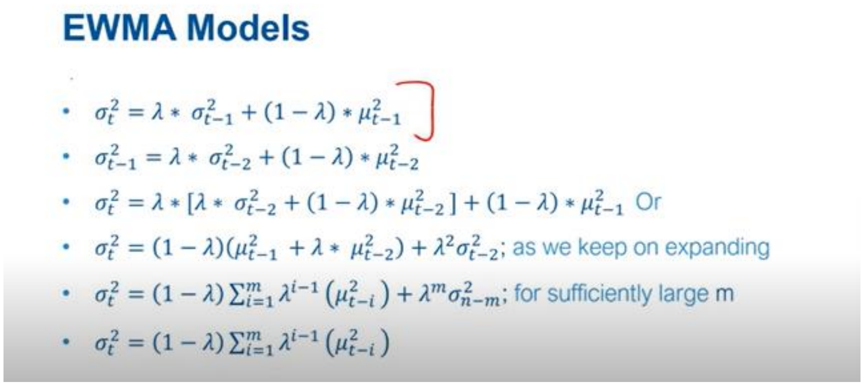


Let us see how this works mathematically. This is the formula for EWMA model which gives sigma square t equal to lambda into sigma square t minus one plus one minus lambda into mu square t minus one

$$\sigma_t^2 = \lambda * \sigma_{t-1}^2 + (1-\lambda) * \mu_{t-1}^2$$

where sigma square t is the estimate of variance for today using historical information. What kind of information? First lambda which is the exponential decay parameter. Customarily the values are like 0.9 or 0.8 in this range. Sigma square t minus one is the estimate of volatility day before while μ_t^2 is the information or innovation term which is usually extracted from the squared residuals or many times it is directly proxied by the square residuals on time t. Now in this model just imagine the impact of the sigma square t minus one as time passes for example its impact on sigma square t plus two and so on let's say on 50th period sigma square t plus 50. A little bit of visualization and imagination about this model would suggest that as the time passes this factor will grow by exponentially by let's say lambda to the power two lambda to the power three and so on lambda to the power 50.

As long as this value is less than one like 0.9 or 0.8 this value very exponentially decays and almost becomes zero in a very few steps and that is a very desirable property of EWMA models in that they exponentially decay the impact of historical volatility estimates and slightly give more weight to the more recent volatility levels. Let's do this in more mathematically engaged manner. So we started with this formula



where sigma square t was the estimate of volatility today using estimate of volatility yesterday and the information or innovation on return squared terms that arrived yesterday. Now in this model we can also write in terms of t minus one for t minus one day this will become t minus two and this will become t minus two. We can substitute this value here to get lambda and then this multiple which is this sigma lambda into sigma square t minus two plus one minus lambda into mu square t minus one plus the original term of one minus lambda into mu square t minus one.

$$\sigma_t^2 = \lambda * \sigma_{t-1}^2 + (1-\lambda) * \mu_{t-1}^2$$

$$\sigma_{t-1}^2 = \lambda * \sigma_{t-2}^2 + (1-\lambda) * \mu_{t-2}^2$$

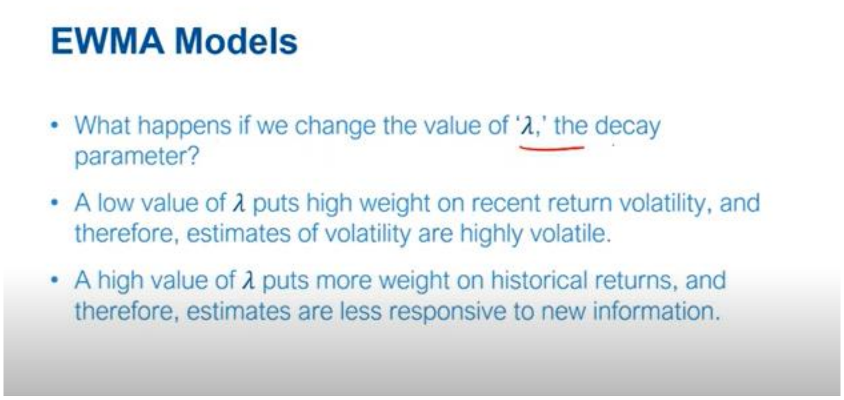
$$\sigma_t^2 = \lambda * [\lambda * \sigma_{t-2}^2 + (1-\lambda) * \mu_{t-2}^2] + (1-\lambda) * \mu_{t-1}^2$$

Now we can take away these error terms or mu square terms together mu square is nothing but the innovation or error or information that has arrived on a particular day. So this becomes one minus lambda into mu square term for t minus one and lambda times mu square t minus two. Now notice as we keep on going further historically in time it will start multiplying with lambda. So for example the mu square t minus three term would be like lambda square into mu square t minus three. Then next term would be lambda cube mu square t minus four and so on and as we keep on going further historically in time for example lambda times n minus one mu square t minus n.

$$\sigma_t^2 = (1-\lambda)(\mu_{t-1}^2 + \lambda * \mu_{t-2}^2) + \lambda^2 \sigma_{t-2}^2$$

$$\sigma_t^2 = (1-\lambda) \sum_{i=1}^{m} \lambda^{i-1} \left(\mu_{t-i}^2\right) + \lambda^m \, \sigma_{n-m}^2$$

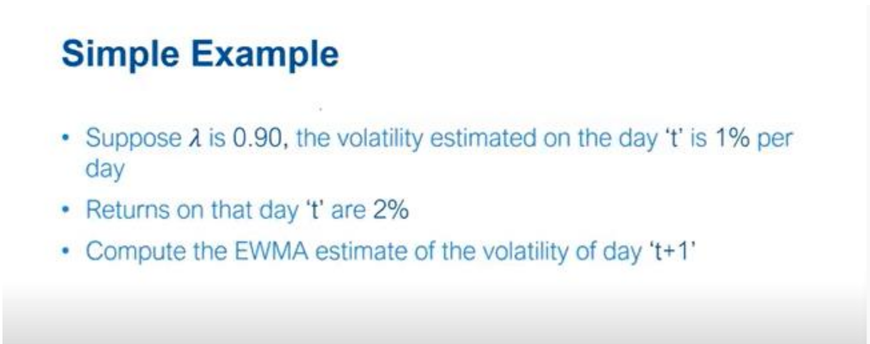
So this way this expansion can be done as we go on historically in past and this observation as we have seen it will keep on moving further in time t minus three and so on lambda square sigma square t minus n as we keep moving historically in time. So a generic term like this would be obtained where the innovation term has this summation series one minus lambda summation i equal to one to m lambda to the power i minus one while the estimate historical estimate of the sigma has lambda to the power m into sigma square n minus m kind of term and if and if this is sufficiently large for example for sufficiently large values of m this term will converge to zero because lambda to the power m where it tends to infinity is equal to zero. So this term will converge to zero and we are only left with this generic term. So this is a more generic formula for EWA model while the previously what we started with this this one is a more basic and easy to interpret formula of EWA which gives that certain weight decay parameter assigned to estimate of volatility sigma square t minus one and one minus lambda assigned to mu square t minus one which is the information or innovation or volatility arrived today or rather saying today it's t minus one. So depending upon whatever I'm forecasting it's immediate previous period.



So looking at this model what happens if we change the value of lambda the decay parameter a very low value of lambda puts a high weight on recent return volatility and therefore estimates of volatility are volatile. So for example if you choose a very sort of low value in this formula sigma t square equal to lambda times sigma square t minus one plus one minus lambda into mu square t minus one if you choose a very low value close to zero then this term will go away and all the weight is assigned to the latest information or innovation or volatility residuals arrived mu square t minus one. So all the weight is given here. However while this is this estimate would be very informative in terms of its recency but it will be very fluctuating it does not account for historical levels of volatility. In contrast if you take a very high level of lambda let's say close to one then almost negligible weights will be assigned to the recent levels while more weight assigned to the historical levels.

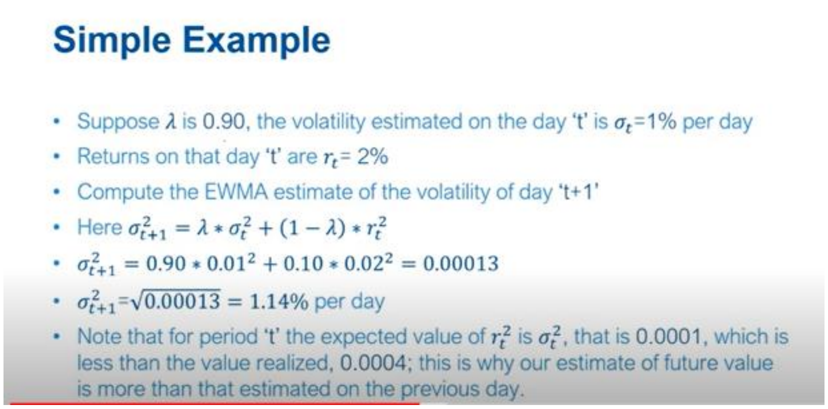
So that is sort of trade-off between recency and reliability of the volatility estimate. To summarize in this video we discussed EWMA models of volatility. We said that these models are driven by the fact that previously historical models they do not account for the fact that more recent levels of volatility should be assigned a higher weight while computing the volatility estimate. So this leads us to either the search of more efficient volatility models leads us to EWMA models which give not only give more weight to recent observations but also the weight of historical observation does not decay in a very sort of radical or one zero kind of manner it gradually or exponentially decays over time. So it is not that till the time the observation is part of sample its entire impact is there and as soon as it goes out of the sample or excluded from the sample it has zero impact. So this was the working of historical volatility model in a EWMA model there is a gradual sort of decay. So when the observation is excluded from the sample it is not that it is having a very sharp impact on the volatility estimate it will not happen with the EWMA but it will happen with the historical model.

In this video we will try to understand the mathematical workings and theoretical interpenings behind EWMA model with the help of a simple example.



Consider the following information the decay factor lambda is 0.9 the volatility estimated on day t that is sigma t is 1 percent the returns on day t that is rt or volatility on day t or mu t observed rather we can call it observed volatility or volatility innovations mu t or r t square under root square t or μ_t we can call this as 2 percent or volatility observed on given day or rather realized volatility you can also call it realized volatility or observed volatility on given day t is 2 percent.

Now we want to compute the EWMA exponential being moving average estimate of the volatility on t plus 1. Let us see how this works. As per the information given to us lambda is 0.9 sigma t is 1 percent and this is daily for daily period and return on day t is 2 percent.



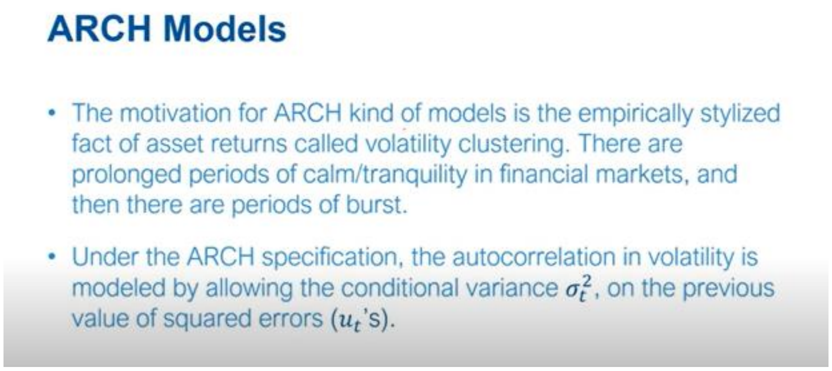
So from this we can estimate we can proxy this to mu t as 2 percent. Now using our formula sigma square t plus 1 which is equal to lambda times sigma square t plus 1 minus lambda into r square t. Now we know the value of lambda 0.9 we also know the value of sigma square t which is square of this which is 0.01 square then 1 minus lambda is this and r t square is 0.02 square. So r t square is being used to proxy the volatility innovations on mu t square. So we get this value as 0.00013. We can take the square root of this to get sigma t as.

Here sigma t plus 1 which is the square root of sigma square t plus 1 as 1.14 percent per day. So now note that for period t the expected value of r square t is sigma square t so expectation of r square t or the volatility on tth day was 0.0001 or 1 percent which was lower than the actual value. So the realized or actual value was 2 percent that is in terms of volatility 0.0004. What does this convey to us? So for example because of this increase so from the historically observed the lower level of 1 percent now the volatility has increased by a shock of 2 percent. So this r t is like a shock of 2 percent. This is sort of positive shock of 2 percent. The volatility on a given day is higher at 0.0004 or sort of in percentage terms 2 percent mu t is 2 percent and this is where estimated future value of volatility is more than estimated on the previous day more as in it is more than this 1 percent so previous day estimate was 1 percent but now it has increased by 0.14 percent to become 1.14 percent because the current day latest shock on volatility was 2 percent which was on the positive side slightly higher. So that is why this estimate has increased. To summarize in this video we saw how to compute EWMA measure of volatility for a simple numerical example. We saw that depending upon the latest more information the latest information show whether positive or negative the recent estimate of volatility changes the recent estimate of volatility changes depending upon more recent observations. Now it also depends on the decay factor lambda how much weight we assign to these recent observations.

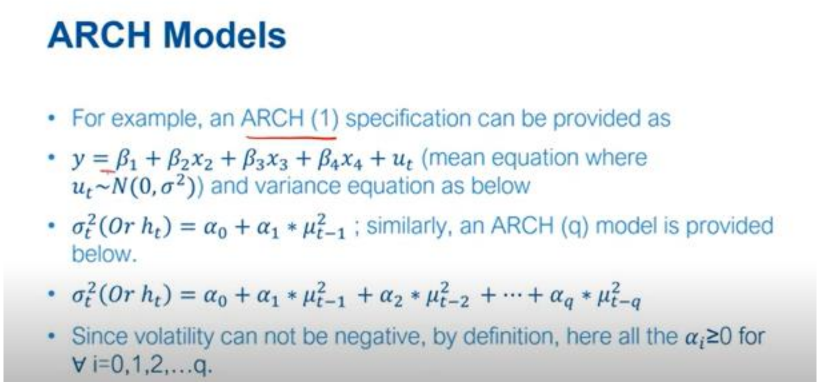
For example if this decay factor is zero for example if this decay factor is very close to zero very small then almost all the weight is assigned to the recent observations if this decay factor is close to zero. However if this decay factor is very large decay factor tends to one then almost all of the weight goes to the historical volatility estimates and these volatility estimates essentially depend on the previous level of volatility levels which were must historically further apart while more recent observations get a weight of 1 minus lambda so if lambda value is very high these recent observations do not get a very high weight and therefore the selection of value of lambda it is sort of compromise between sort of trade off between more recent observation or more reliability. So if you give more weight to more recent observation that is RT square your estimate will be very noisy very fluctuating but if you get very less value to it your estimate would be less timely it will be sort of average of historical values of historically estimated values and therefore not reflect a very timely signal to you of volatility.

Previously with the class of EWMA models we captured a very simple yet powerful idea that all the observations all the innovations should not get equal weight while estimating volatility and more recent observations should get a higher weight. A natural offshoot and a more systematic and robust offshoot of this concept or the idea conveyed by EWMA model is ARCH family of model which capture this fact that volatilities auto correlated or auto regressive in a very systematically and mathematically more robust manner.

In this video we will introduce the ARCH family of models and their mathematical modeling.



The motivation behind ARCH family of models is the empirically stylized fact of a certain returns called volatility clustering that is there are periods of prolonged periods of calm and tranquility in financial markets and then there are periods of bust and periods of calm and tranquility occur together while periods of bust also occur together. This is called volatility clustering under the auto regressive conditional heteroscedasticity or ARCH specification this auto correlation is volatility is modeled by allowing the conditional variance and we have already discussed what is the conditional variance σ_t^2on the previous value of squared errors or mu t s or what we call as innovations or information shocks that are employed to model the conditional variance orσ_t^2.



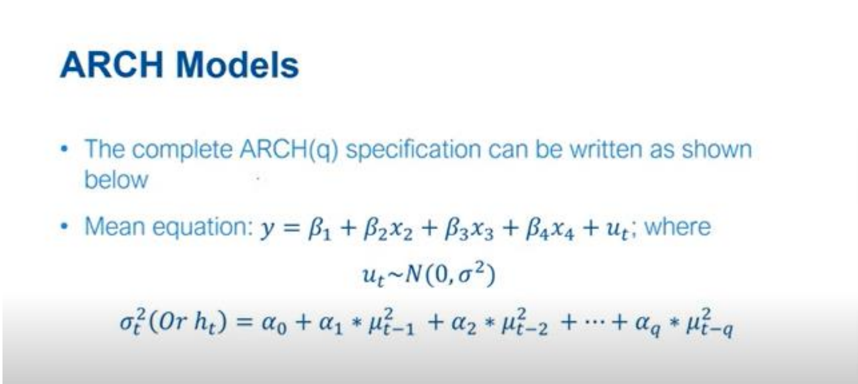
In the ARCH family of models let us start with what we call as ARCH(1) specification. In general the arch model or any volatility model is a combination of mean or sort of return model which appears like this a linear function of y return y here would be return some function of some exogenous variables and some lags of returns a linear function which appears like this and an error term which is the sort of information or innovation shock or information that has arrived in the period t.

So this is what we call as mean equation where this shock information shock is modeled as mean 0 with the constant variation sigma square. So this error term is modeled mu t as a normal distribution with a mean of 0 and standard deviation of sigma this is the distribution of error and then we have the conditional variance equation as sigma square t which is a constant term alpha naught plus alpha 1 into mu square t minus 1 where these innovation or error term square mu square t minus 1 are obtained from the mean model.

$$\sigma_t^2 = \alpha_o + \alpha_1 * \mu_{t-1}^2 + \alpha_2 * \mu_{t-2}^2 + \cdots + \alpha_q * \mu_{t-q}^2$$

Please note the focus is not here on mean model which is a basic requirement basic mandatory requirement to obtain these mu t's here we model the return with some of its own lag terms plus some exogenous variables that are predictor of returns and then extract the residuals that is mu t which sort of reflect the current information arrival the information shock that has arrived in the current period and now this mu t will be used to model the volatility conditional variance through this formula. This is a simple ARCH (1) estimate this is a simple ARCH(1) specification where estimate of conditional variance or sigma square t is a function of previous period innovations that is mu square t minus 1. I repeat again this is a simple ARCH (1) specification where estimate of volatility is a function of its immediate previous period innovation or information shock or residuals that is mu square t minus 1 this is ARCH(1) specification.

In similar manner we can think of Arch (Q) specification as sigma square t which is dependent on previous Q periods innovations or residuals or information shock in squared form. So many times when residuals are not available we tend to use returns to proxy these as duals directly for example I could use to proxy mu square t minus 1 I could use R t minus 1 square or mu square t minus 2 as R square t minus 2 this we have done in the previous numerical example if you recall so this can be done that and the distribution of mu is like this it is mean of 0 with a sigma square variance. Now here one thing important to be noted because the sigma square t is also positive and these mu square t are all positive the coefficients and like alpha naught alpha 1 alpha 2 they cannot be negative and therefore some constraint has to be put so that all the alphas are greater than 0 for all i equal to 0, 1, 2 and up to Q. So all these alphas have to be greater than 0 which is a sort of sufficient condition to model this.

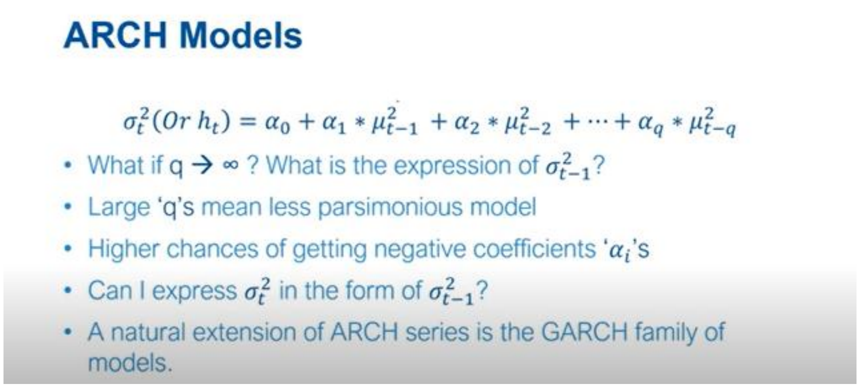


So the resulting Arch Q a more generic version of Arch model that is Arch Q model specification can be written as a combination of mean equation which is y equal to beta 1 plus beta 2 x 2 plus beta 3 x 3 and so on plus mu t where mu t is normally distributed with a mean of 0 and a variance of sigma square and then the conditional variance expression or the conditional variance estimate of Arch model sigma square t for period t is written as either H t you can also write as H t or sigma square t as alpha naught plus alpha 1 plus mu square t minus 1 plus alpha 2 into mu square t minus 2 and so on till alpha Q into mu square t minus Q where all these mu squares are the innovation or error terms for models for period t minus 1, t minus 2 and up till t minus Q.

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \mu_t, where\ \mu_t \sim N(0, \sigma^2)$$

$$\sigma_t^2 (or\ h_t) = \alpha_0 + \alpha_1 * \mu_{t-1}^2 + \alpha_2 * \mu_{t-2}^2 \dots \dots + \alpha_q * \mu_{t-q}^2$$

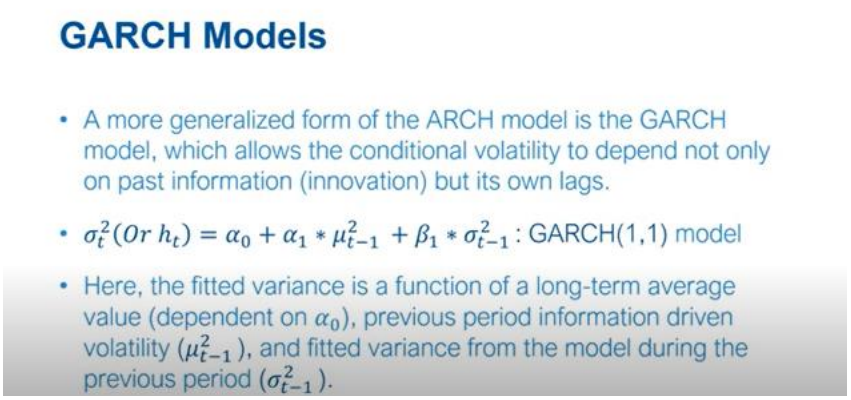
And please remember one external constraint that all these alpha naught and alpha 1 and alpha 2 have to be greater than equal to 0 that has to be put to ensure that in none of the estimates the estimate turns out to be negative which will be the next previous estimation.



However, this model has particularly two problems one is that what if the lag structure is very long then this model becomes as you can see very extremely un-parsimonious if lag structure is long. Second, you would a simple way to solve that kind of issue and you can see that here would be to somehow get the expression in the form of sigma square t minus 1 which as we will see the Garch family of models is a very simple and intuitive extension to this form of ARCH models. The idea here is that large Q s mean that models very less parsimonious and there is a higher probability of getting negative coefficients or alpha i s which leads us to think of a simpler ways and of expressing this model and as you would have already thought that probably and the intuition comes from EWMA model that probably we could replace some of these mu i s and express this sigma square t estimate in terms of previous period estimates of sigma square t minus 1 because essentially the information that is captured in sigma square t some of that is also capturing captured in sigma square t minus 1 in a slightly more parsimonious manner. So, rather introducing all such long like the structure we could rather in a parsimonious manner use somehow sigma square t minus 1 and that leads to the natural expansion of Arch series towards Garch family of models.

To summarize this video, we noted that the Arch series of models very nicely capture the idea that the volatility estimates should have higher weight in the more recent terms and as the time passes and a particular period becomes older and older its impact on the volatilities estimate should decay. However, the Arch family of model has its own set of problems such as it has slightly less parsimonious structure and we also have to externally put constraints so that coefficients remain positive all the coefficients remain positive. To account for these problems a natural offshoot is a Garch family of models which is essentially derived from Arch family of models which we will discuss in the next set of videos. In the previous discussions we noted that even though Arch set of family models are extremely useful in modeling volatility though they still have their own set of problems. For example, an extremely non-parsimonious structure of lags or lagged volatility levels makes it vulnerable to negative coefficients or resulting in negative coefficients in estimation.

In this video we will introduce and in a series of next few videos we will introduce Garch family of models and how they improve upon the Arch modeling.

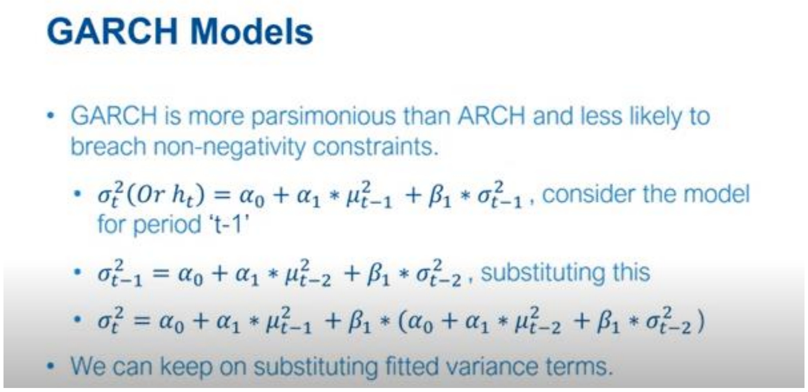


In a sense the GARCH models or generalized what we call generalized Arch models they are essentially the generalized version of Arch model that allow the conditional volatility sigma square t to depend not only on the past information or innovation terms that is mu t but also its own lags that is sigma square t minus 1 and previous lags. Let us see how. So, our estimate of volatility which is sigma square t or H t has three important components alpha naught, alpha 1 into mu square t minus 1 plus beta 1 into mu square t minus 1 this is called simple Garch 1 specification and as we will see shortly it is a very powerful and useful specification.

$$\sigma_t^2 = \alpha_0 + \alpha_1 * \mu_{t-1}^2 + \beta_1 * \sigma_{t-1}^2$$

First and foremmeansthe term alpha naught this term alpha naught makes this model capable to handle what we call as long term mean reversion property that means because of this term this term helps model achieve some kind of long term means so that when the value is higher the sigma square t estimate is higher than the normal levels it tends to pull it towards some kind of long mean or whether it is lower then also it pulls higher towards that long term mean.

So, this alpha naught drives that property of long term mean reversion we will see the value of that long term mean shortly. Now this mu square t minus 1 into alpha 1 creates that dependence on recent information. So, this mu square t minus 1 captures the information or gives sort of alpha 1 way to the latest information that has arrived and all the previous levels of information are captured through the sigma square t minus 1 with the idea that historical information structure will be captured through this sigma square t minus 1 and its coefficient beta 1. Now because of this extremely, extremely parsimonious structure the coefficients alpha 1 alpha naught and beta 1 only three coefficients being present the chances of these coefficient turning to be negative is very less.

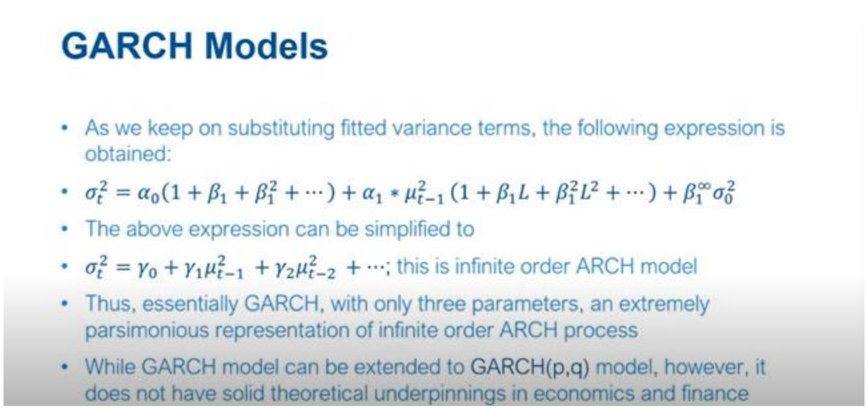


Let us see when we make the statement that GARCH is a more parsimonious model as compared to ARCH and very less likely to reach non-negative constraints.

Let us see why we could make this statement. If this is the model if the generalized is called GARCH 1 1 model so you have 1 lakh for mu square t minus 1 and 1 lakh for the previous conditional estimate sigma square t minus 1. So, your volatility estimate is dependent on previous values of mu square t minus 1 and sigma square t minus 1. So, it becomes this kind of model. Now you can simply substitute for sigma square t minus 1 in the same manner to get it in the form of mu square t minus 2 and sigma square t minus 2. These values can be substituted further here resulting in this kind of model and as you would have now guessed it we can keep on substituting fitted variance terms like this.

$$\sigma_t^2 = \alpha_0 + \alpha_1 * \mu_{t-1}^2 + \beta_1 * (\alpha_0 + \alpha_1 * \mu_{t-2}^2 + \beta_1 * \sigma_{t-2}^2)$$

So, this is our basically fitted terms when I use the word fitted variance this is nothing but sigma square t or its different variance.



So, as we go on substituting these fitted variance terms, we can get this expression.

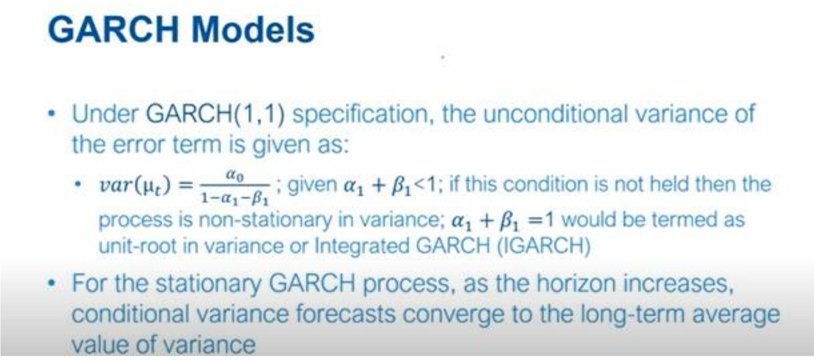
$$\sigma_t^2 = \alpha_0(1 + \beta_1 + \beta_1^2 + \cdots..) + \alpha_1 * \mu_{t-1}^2(1 + \beta_1 L + \beta_1^2 L^2 + \cdots.) + \beta_1^{\infty} \alpha_0^2$$

We keep on iterating in terms of mu square t minus 1 and taking the sigma square back in time for example sigma square t minus 1 to sigma square t minus 2 and so on up till infinitely long ahead in time and therefore this term and generally it is assumed that all these coefficients are less than 1 so this term will approach to 0. So, essentially you would get a term which is a some kind of constant term and then another terms which are function of lags of mu square t minus 1 for example this would be mu square t minus 1 and this L represents lag for example this will be mu square t minus 2 this will be mu square t minus 3 and so on. So, essentially you can think of this expression as a some kind of arch model like this this is some kind of arch model with a constant term and lags of mu square t minus 1 but because we have taken or absorbed so many lags we can think of it as infinite order arch model.

$$\sigma_t^2 = \gamma_0 + \gamma_1 \mu_{t-1}^2 + \gamma_2 \mu_{t-2}^2 + \cdots \dots \dots ..$$

So, this is like an infinite order arch model and thus essentially with only three parameters that is alpha naught alpha 1 and beta 1 we could express an infinite order arch process in a simple GARCH 1 process and this is the very reason that while this model can be extended as GARCH PQ as well where you have P order of mu square t minus 1 that is mu square t minus P and sigma square t minus Q but in general there is no theoretical underpinning for this kind of model in economics and finance and a GARCH 1-1 kind of model is more than capable enough to model most of the series in economics and finance and that is why because of this extremely parsimonious nature the chances of these coefficients turning to be negative is very less and it provides a considerable improvement over the arch family of models because it captures infinite series of arch model with just simple GARCH 1-1 specification generalized arch. To summarize in this video we introduced the GARCH model as a simple combination of three terms which is some kind of long term mean alpha naught a coefficient alpha 1 which is assigned to innovation terms mu square t minus 1 and alpha 2 assigned to historical estimates of sigma square t minus 1 with such a simple and parsimonious structure it could capture a model which is equivalent to an infinite arch process so an infinite arch process is simply captured by this kind of GARCH 1 model although we could generalize it but in economics and finance this GARCH 1 model is reasonably capable and robust to handle any kind of time or price series and therefore in this model the chances of coefficient turning to be negative is also very less.

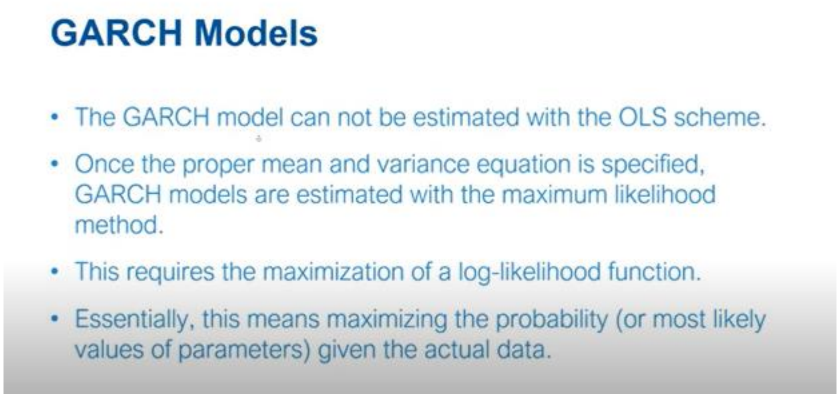
In this video we will discuss the GARCH 1-1 model, which is a very powerful and useful model in more detail. We will also discuss some of the issues with GARCH 1 model that leads to search or requirement of more advanced GARCH models. Recall our expression for GARCH 1-1 model as sigma square t which is a combination of constant alpha naught plus alpha 1 into mu square t minus 1 plus beta 1 into sigma square t minus 1.



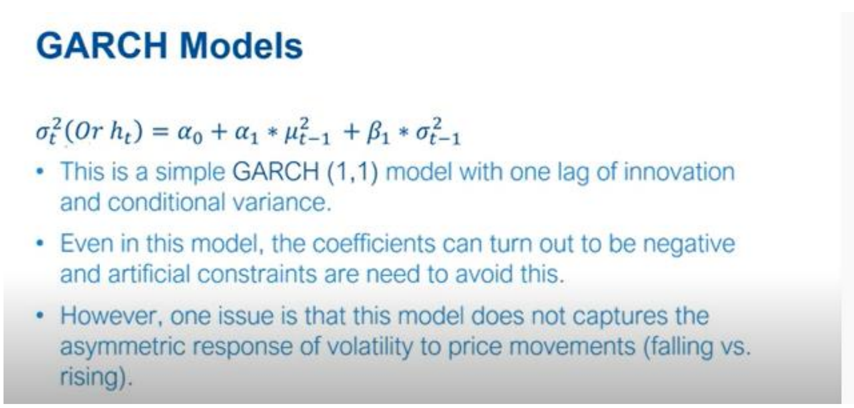
Here the unconditional variance of the error term is given as alpha naught upon 1 minus alpha 1 minus beta 1. One required condition is that the summation of alpha 1 plus beta 1 is less than 1. If this condition is not held then the process is non-stationary in variance probably many of us would have heard this term non-stationary mean or stationarity of a series this is what we call as non-stationary or stationarity of variance. So if alpha 1 plus beta 1 is less than 1 then the process stationary but if this condition is not held then it is non-stationary. While there is no precedence or no rational for alpha 1 plus beta 1 greater than 1 which essentially would mean that the volatility explodes, volatility sort of explodes.

This is called explosion of volatility infinitely in times to come but generally there is no precedence of that at best what you have is alpha 1 plus beta 1 equal to 1 which is often termed as unit root or non-stationarity in variance or integrated I-GARCH. So this relationship of alpha 1 plus beta 1 equal to 1 ($\alpha_1 + \beta_1 = 1$)captures what we call as non-stationarity or unit root process in I-GARCH integrated actual variance. So given that a GARCH process stationary this unconditional variance what is the application of this unconditional variance if the GARCH process stationary as you keep on forecasting the future values of sigma square t let us say sigma square t plus 1 sigma square t plus 2 sort of unconditional forecast. So you keep on forecasting as horizon increases from 1, 2, 3, 4, 5 and days after certain time the historical information that you had till time t equal to 1 the impact of that information will die away and the volatility estimates will converge to this value. So what we obtain here is a very useful property what we call as long term mean reversion long term mean reversion what it means is that if the GARCH process stationary as the horizon increases and you keep on forecasting for t equal to 1, t equal to 2, t equal to 3 and so on the impact of historical information that was available at t equal to 0 dies and what you are left with is sort of unconditional forecast which converge to their long term value long term unconditional value which is provided here this is a very useful property of

GARCH 1 specification.



Now this GARCH model as you would have noticed nonlinear in nature so it cannot be estimated with ordinary least squares scheme. Once you have the idea of proper mean equation and variance equation as we have already discussed some variants of mean and variance equation so once you have some idea of these mean and variance equations you need to estimate them with what you call as maximum likelihood method the discussion the elaborate discussion of maximum likelihood estimation is out of the syllabus and not part of this discussion. So we said that GARCH model effectively requires maximization of some kind of log likelihood function basically this essentially it means obtaining some sort of parameters a family of parameters for GARCH in this case the parameters are alpha naught alpha 1 and beta 1 as we saw that maximize the probability of getting the observed or actual data so these parameters observing those parameters that maximize the probability of getting the actual data that we observe from financial markets so that would be the scheme of MLE or what we call as maximum likelihood estimates these estimates of alpha naught alpha 1 and beta 1 would be employed. Sometimes the model or the process of MLE does not exactly converge if the observations are too less than it may not converge and various other reasons it may not converge as well it is a non-linear kind of iteration where you maximize some kind of probability density of getting the observed data.



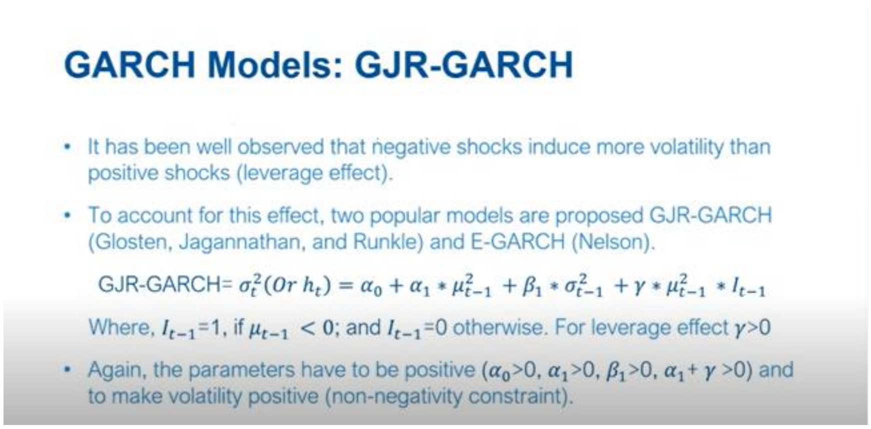
So now to summarize this video we discussed that GARCH 1 1 process appears like this sigma square t equal to alpha naught plus alpha 1 into mu square t minus 1 plus beta 1 into sigma t square t minus 1 which is a GARCH 1 1 model with 1 lag of innovation or error term mu square t minus 1 and 1 lag of conditional variance which is sigma square t minus 1.

$$\sigma_t^2 (or\ h_t) = \alpha_0 + \alpha_1 * \mu_{t-1}^2 + \beta_1 * \sigma_{t-1}^2$$

However, notice in this model even though very parsimonious it still has three coefficients which may turn out to be negative so we need some kind of artificial constraints to ensure that these coefficients alpha naught alpha 1 and beta 1 do not turn out to be negative. The second and also very important this model does not capture the asymmetric response of volatility or what we call leverage effects of price movements. Recall we earlier said that in financial markets what we observe is called leverage effect where rising prices result in sort of lower levels of volatility while falling prices results in higher level of volatilities or price innovations or shocks to price that are on the positive side have less impact on volatility while those that are negative have a higher impact of volatility. So this asymmetric nature or asymmetric behavior of volatility is not observed or modeled by this kind of model.

There is nothing that captures or models this asymmetry in volatility behavior. So these are some of the shortcomings of this model and in the next video we will discuss some of the more advanced models that tries to overcome these handicaps or shortcomings of the GARCH model.

In this video we will conclude our discussion of GARCH family models by introducing two models that is GJR model and eGARCH model that tries to overcome the shortcomings one that is leverage effect, help us model leverage effect and second negativity constraints of model coefficients.

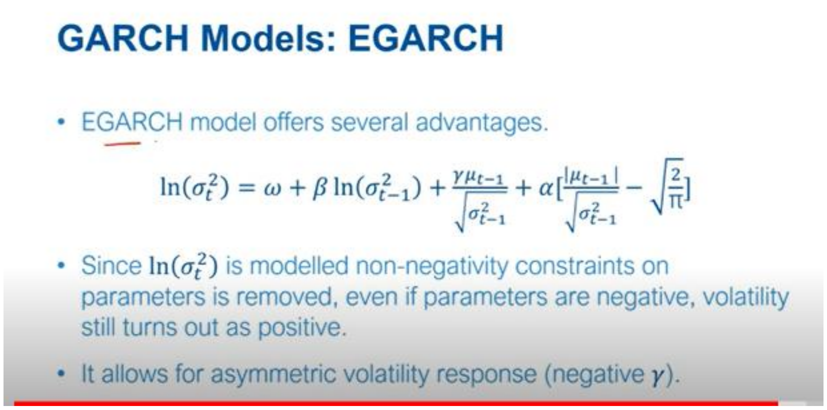


So we start the discussion with the leverage effect and it has been well observed that negative shocks induce more volatility than positive shocks as we discussed earlier and we gave it a name leverage effect. Now there are two very important models, one is called GJR GARCH model and eGARCH, Nelson and Seagarch model that tries to overcome this. Let us start with the GJR GARCH model which appears like this where estimate of volatility sigma square t is a function of alpha naught, alpha 1 and beta 1 which are familiar terms with us.

$\sigma_t^2 (or\ h_t) = \alpha_0 + \alpha_1 * \mu_{t-1}^2 + \beta_1 * \sigma_{t-1}^2 + \gamma * \mu_{t-1}^2 * I_{t-1}$We already saw them in GARCH 1 1 but one more term is added which is gamma into mu square t minus 1 into i t minus 1. So we already understand these terms GARCH 1 1 parameters. Let us discuss this particular last term. Here i t minus 1 is sort of indicator or dummy variable which takes a value of 1 if gamma mu t minus 1 is less than 0 or it is equal to 0 otherwise. So essentially we are running two models, so essentially we are running two models one when the shock is positive, when the shock is positive then only this version GARCH 1 1 is run but when the shock is negative then we are running a slightly more elaborate model to capture that asymmetric response.

The idea here is that with running this asymmetric or sort of different models for positive and negative shock that the negative shock, the negative original shock would be captured by this gamma and as you would have guessed if the leverage effect are indeed significant, if the leverage effect are indeed significant then this gamma which captures the leverage effect should be significantly positive. So for the leverage effects to exist this gamma should be greater than 0 because this i t minus 1 is 1 only for negative shocks. For the positive shocks it is 0 so only this model is 1. So, if there is some incremental positive or higher impact of negative shocks that would be captured by the positive coefficient through this complete term, positive coefficient gamma and this complete term. So, again this model has parameters alpha naught, alpha 1, beta 1 that were only from GARCH 1 1 and then additional gamma.

Now for this model also the parameters that is alpha 1 naught, alpha 1, beta 1 and alpha 1 plus gamma have to be positive or greater than 0 to ensure that the estimate of volatility sigma square t is positive. So, that additional constraint has to be put although it accounts for the asymmetric nature of volatility or leverage effect, but still non-negativity constraint has to be put externally.



Another very important model of Nelson-Egache model captures both the properties that is non-negativity and as well as the leverage effect.

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \frac{\gamma \mu_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha[\frac{|\mu_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}]$$

Look at the expression here you have estimate in the natural log form then the omega beta into natural log of sigma square t minus 1 and then mu t minus 1 into gamma upon square root sigma square t minus 1 plus alpha and magnitude of the error terms mu t minus 1 upon sigma square t minus 1 and so on. In this model notice that estimate of volatility is modeled in the form of natural log and therefore the non-negativity constraint is removed because whatever even though the parameters can be negative the estimate of volatility may still be positive because of this log thing.

So, this model affords it can afford the negative parameters that is a very useful property and volatility estimate will still turn out or rather conditional volatility estimate will still turn out positive. The second and very important property of this model is the asymmetric response of volatility while the mod of mu t minus 1 captures the symmetric response or the both sides it considers negative and positive sides as same, this mu t-1 captures the asymmetric nature. Now as per out knowledge of asymmetry effect negative shocks should have high impact on volatility and therefore we are estimating the gamma to be negative and significant , which would capture the volatility shock in this volatility model.

To summarize this video, we studied two advanced GARCH models, GJR-GARCH and E-GARCH. GJR which could account for asymmetry or leveraging effect but it still has the problem of externally put non-negativity constraint that is non-negativity constraint in parameters or coefficients , i.e., alpha not, alpha one, beta one has to be externally put. Whereas nelson EGARCH model can overcome both these constraints that it could not only model the leverage effect but also we need not to put any non-negativity constraint because of natural log to estimate even though coefficient can be negative, the estimate of volatility will still be positive.

shocks induce more volatility than positive shocks as we discussed earlier and we gave it a name leverage effect. Now there are two very important models one is called GJR GARCH model and Nelson EGARCH model that tries to overcome this. Let us start with the GJR GARCH model which appears like this where estimate of volatility sigma square t is a function of alpha naught, alpha 1 and beta 1 which are familiar terms with us.

We already saw them in GARCH 1 1 but one more term is added which is gamma into mu square t minus 1 into i t minus 1. So we already understand these terms GARCH 1 1 parameters. Let us discuss this particular last term. Here i t minus 1 is sort of indicator or dummy variable which takes a value of 1 if gamma mu t minus 1 is less than 0 or it is equal to 0 otherwise. So essentially we are running two models, so essentially we are running two models one when the shock is positive, when the shock is positive then only this version GARCH 1 1 is run but when the shock is negative then we are running a slightly more elaborate model to capture that asymmetric response. The idea here is that with running this asymmetric or sort of different models for positive and negative shock that the negative shock, the negative original shock would be captured by this gamma and as you would have guessed if the leverage effect are indeed significant, if the leverage effect are indeed significant then this gamma which captures the leverage effect should be significantly positive.