Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 10

Lecture 46: Stochastic Volatility Modelling

Hello all, welcome to this course on time series modeling and forecasting using R. Now, this is a fresh week and a fresh new topic, and of course, you can see the topic in front of you. So, the broad idea for this entire week would be volatility modeling, or stochastic volatility modeling, and in general, we will refer to it as stochastic volatility modeling. Now, first, we will break each and every point into different aspects. By the way, this entire week will be slightly more applied because, let us say, if you take the area of finance, right? So, let us say if you have some interest in, let us say, stock price forecasting or, let us say, order book dataset forecasting. Or, let us say, mid-price forecasting.

So, all these things contain some volatile aspects in them. So, first, we will understand what you mean by volatility. And then we will break it down into the impacts of volatility, different kinds of volatility, and eventually, we will introduce some time series models specifically for analyzing and modeling the volatility nature of a time series. Hence, the name stochastic volatility modeling. So, rather than the actual time series values changing or the mean changing, what would happen if the variance also changes, by the way, right?

So, the entire stochastic volatility modeling focuses more on the changing variance aspect of a time series, okay? Alright. So, again, we will see a lot of examples; do not worry. So, since this is just the first session this week, we will try to provide some introduction, give some examples down the line, and so on, okay? Okay, so the first slide is a bit of an introduction.

So, so far what we have seen is we have talked heavily about the ARIMA model. So, what is ARIMA? So, Auto Regressive Integrated Moving Average Models. And all ARIMA models have a tendency to describe the mean development of a series, right? So, if you pick up, let us say, even AR models, MA models, ARMA models, or in general ARIMA models, they tend to focus only on the mean of the series and not the variance, isn't it?

So, when we try to model any practical time series data using an ARIMA structure or, let us say, for that matter, a SARIMA structure. So, both ARIMA and SARIMA only focus on the mean aspect of the series. This is the first point. Now, the second point is optimal forecast. So, the conditional mean of the model, which minimizes the forecast error variance, was taken to be the optimal forecast, right?

So, let us say when you put forward an ARIMA model and then try to forecast down the line. So, how did you try to figure out the best forecast? So, the best forecast or the optimal forecast was the one where the conditional mean of the model minimized the forecast error variance, right? Now, on the other hand, one can actually talk about something called unconditional variances or time invariance. So, when you focus on unconditional variances, now again, do not worry.

So, in the subsequent slides, I will also elaborate more on what do you mean by unconditional variances, what do you mean by conditional variances, etc., So, the third point is unconditional variances are time invariant. So, if you have a certain variance which is unconditional which does not depend on anything. So, such variances can be assumed to be time invariant. So, time invariant means they do not change with time basically.

But the problem lies the moment you talk about conditional variances. So, conditional variances keep on changing right. So, the conditional variances and covariances often depend on the magnitude of past shocks and past variabilities. So, here the third point again just to summarize very quickly I will try to break it down in two parts. So, you have two different kinds of variances.

First one is called as unconditional where you do not have any restrictions and hence the variances are not based on time or rather, they are time invariant. And the second kind of variances are called as conditional variances and covariances which often depend on the magnitude of past shocks and past variability. So, such conditional variances have a tendency of depending on its own lags, right? Or rather its own pass variability, etc. And then the last point is in modern economic theory, increasing role of risk and uncertainty considerations makes the modeling of time varying variances and covariances crucial, right?

So these days, you must have seen a lot of examples where suddenly something happens, some news comes in, and then the price sort of fluctuates very drastically. So all these aspects are kind of changing variance aspects, right? Just to give you a very brief graph as a starter, let us say if you think of a stock price which moves in this fashion, but suddenly due to some unforeseen news, the price fluctuates very drastically, and here you can clearly see that you have some changing variance problems in the series now. So, eventually, you have some highly fluctuating prices or rather returns, let us say, due to some external factors. Overall, if you observe this entire series, then you can see that you have some variance issues here.

So, the variance is not constant here. And hence, we cannot model such series using simple ARIMA models and so on and so forth. We require to model the underlying volatility as well. So, the last point again, just to summarize, is that in modern economic theory, the increasing role of risk and uncertainty considerations makes the modeling of time-varying variances and covariances very crucial. So, I think this slide gives you slight motivation as to why we should even study volatility or why we should put forward models to capture volatility, etc.

So, before you move forward with any of the modeling aspects, I will show you some very quick examples. So, the first example contains a chart of VIX. So, we have something called the VIX index. So, the VIX index is nothing but the volatility index. So, the full form of VIX is nothing but the volatility index, and again, probably I am very sure that many of you might already have an idea as to what you mean by VIX.

So, VIX stands for how volatile the index is, or it sort of captures the overall variance of the index. Okay. So, by the way, each and every line or each and every bar that you see here does not signify any price per se. So, this is not a price plot. This is rather a plot of the underlying variance of the index, and hence it is called the volatility index.

So, how does the index behave from a volatility point of view? Okay. Now again, here, clearly the data is from January of 2010 all the way up to, let us say, December of last year, which is 2023. And here, clearly, you see that whenever you see peaks here, those are kind of thresholds, and those are periods of high volatility. And on the Y-axis, you can see the actual value of the VIX.

So, for example, here, this is exactly the COVID time. So, you see that the markets were highly volatile. And since the markets were highly volatile, the VIX shot up, and the value of the VIX happened to be around 80 here. And again, here you can see some kind of high peaks. So, let us say the first peak would be somewhere here, and the second peak is here.

So, at all these time points, there was some activity in the market, which made the market highly volatile. So, again, just to summarize, WIX is nothing but a graph that captures the underlying volatility of an asset. So, in this case, it is an index. And now, the second example could be, let us say, the daily returns of the NYSE. So, the New York Stock Exchange Composite Index.

So, this is exactly how the NYSE returns have been from 2018 to very recently. So, in this year itself. So, again here, you can clearly see that around COVID, since the NYSE was, in fact, very volatile. So, you see sudden disruptions in the entire behavior. So, you see a lot of fluctuations in probably this area or again slightly here, right?

So, all these large fluctuations make the variance not constant, right? So, overall, if you observe the variance here. So, the variance is obviously not constant. So, you have some constant part here, but due to these sudden fluctuations, the variance also goes for a toss, right? So, again, this is a small disclaimer that when I say these two things, they sort of mean the same thing: changing variance in a real-time series and the time series being volatile.

So, if a time series is volatile, what it means is the variance keeps changing with time. So, the sigma square that you see—sigma square itself—is a function of time. So, you can actually say sigma square with subscript t. So, again, if variance keeps changing with time, we will say that the underlying time series is highly volatile. Now, the next thing is: what are the properties of volatility? So, the first one is that volatility clusters exist.

So, in finance literature, we have this important idea called volatility clusters. So, what do you mean by volatility clusters? So, volatility is high for certain time periods and low for other periods. So, let us say volatility—I will show you through a graph. By the way, volatility clusters mean that if you see high volatility,

that period is almost always followed by high volatility. And whenever you have low volatility, that period is again followed by low volatility. So, initially in the series, you see that the series is highly volatile, and down the line, you see that the series is not that volatile. So, high or low volatility is followed by the same kind of volatility structure. So, high volatility is followed by high, and low volatility is followed by low.

So, such periods are called as volatility clusters. So, volatility often exhibits a clustering kind of a mechanism. Second property is volatility evolves over time in a continuous manner and that is volatility jumps are there. So, what do you mean by that? That volatility does not jump usually.

So, you would not see something like you have very highly volatile series and suddenly after a jump the series let us say the variance kind of drops right because you do not see highly volatile nature here, but then such jumps are not very evident. So, it should be in a continuous manner right. So, highly volatile periods and low volatility should be again going across the time scale in a continuous manner. Thirdly, volatility is often stationary and does not diverge to infinity. It sort of varies within fixed limits.

Again, this is a very important point that how do you again remember how do you capture volatility? So you capture volatility by the length or rather the height of all these fluctuations, isn't it? So, of course, variance cannot go to infinity. So, it has to stop somewhere. So, even if you have really really high spikes and really really high volatility clustering, but again the peaks will not go to infinity.

So, hence we say that volatility is often stationary and does not diverge to infinity. It sort of varies within some fixed limits. And last property is that volatility seems to react differently to a big price increase and a big price drop. So, if you see a big price increase vis-a-vis a big price drop then the volatility kind of changes accordingly. So, for any big price increase you will see a different volatility change, for any big price drop you will see a slightly different volatility change.

What do you mean by this is that the big price drops usually have a greater impact. So, whenever the price drops. So, price drops are much more drastic than price increases. So, whenever you have a downward sloping trend in a market or whenever the markets are not doing or rather not performing well. So, the market is going down.

So, during those times you will see highly volatile kind of structures rather than whenever a market is going up ok. So, again the big price drops usually have a greater impact this phenomena is rather referred to as a leverage effect ok. So, this phenomena is referred to as the leverage effect. So, again these are some properties of volatility. So, volatility clusters then how exactly can you specify volatility.

So, volatility evolves in a continuous manner. Volatility can be assumed to be stationary, as it does not diverge to infinity, etc. And lastly, you have this very important leverage effect underlying the volatility. So, I thought of putting a couple of slides on changing variance. So, what do you mean by changing variance?

So, firstly, we will try to understand these two technical terms. Probably, we have revisited these terms even earlier. So, a time series is said to be heteroscedastic. So, the first term is heteroscedastic. So, a time series is said to be heteroscedastic if its variance changes over time.

So, you have to keep this in mind that whenever you see changing variance, the underlying time series is heteroscedastic. As opposed to that, if the variance is not changing and is rather fixed, then we will call such a series homoscedastic. Again, a very quick example could be, let us say you have a series which looks like that. So, this could be an example of homoscedasticity because the variance is kind of constant. But at the same time, if you have a slightly different series which behaves like this, then you see some changing variance problems here.

So, such a series is called heteroscedastic. And the second point is whenever the variance is not constant, it follows a mixture normal distribution. And one can actually expect more outliers than expected from a normal distribution. That is, when a process is heteroscedastic, it will follow a heavy-tailed or outlier-prone probability distribution. So, again, changing variance is a problem.

So, whenever variance keeps on changing, all the normality assumptions, all the distribution assumptions go for a toss. So, whenever the variance keeps on changing, we cannot apply a proper normal distribution to the errors. So, the error distribution would not be normal. So, the distribution of the errors would contain heavier tails than a normal distribution, or you might see some outliers in the distribution, and so on. So, again, heavy-tailed and containing lots of outliers is a problem because the normal distribution is not being followed.

So, again, just to summarize, whenever a time series is heteroscedastic, we end up with some problems like these. So, until the early 80s, econometrics had focused almost solely on modeling the means of the series. So, in the early 80s, they did not worry too much about the changing variance or volatility, etc., that is, their actual values. But very recently, however, researchers have started focusing increasingly on the importance of volatility, its variations, and its effects on the mean values. So, along with studying the mean or modeling the mean, how can you put forward a model to capture the changing variance or the volatility nature of the series as well? So, henceforth, a key distinction is between the conditional and unconditional variance.

So, as discussed even before, you have two different kinds of variances: first is conditional, second is unconditional. So, how do you differentiate between conditional variance and unconditional variance? That is the key here. A short description about what you mean by conditional variance and what you mean by unconditional variance. So, the first thing is unconditional variance. Unconditional variance is any common variance you must have studied in any other stats course before this. You do not have any restrictions here.

$$Var(X) = E[X - E(X)]^2$$

So, unconditional variance is just the standard measure of the variance given by this equation. So, variance is nothing but the expected value of x minus its mean, whole squared. So, expectation of x minus E of x. E of x is nothing but the mean, and then whole squared. But the moment you come down to conditional variance, things are slightly different. So, conditional variance is nothing but the measure of the uncertainty about a variable given a model and some information set.

So, let us say this capital omega is nothing but some information set you have. And then, on top of that, how do you condition based on that information set and then find the variance of that? So, such a variance is called conditional variance. So, again, if you see the formula, it is slightly different. So, expectation of X minus E of X given omega, and then whole squared.

$$Cond\,Var(X) = E[X - E(X|\Omega)]^2$$

So, here in conditional variance, you have this extra omega term, which is nothing but the assumed information set that you already have. Okay, now the next thing is stylized facts of asset returns. So firstly, you should understand what you mean by stylized facts. So often in finance literature, you have a thing called stylized facts. So stylized facts are nothing but some facts of that particular asset class or, let's say, prices or returns, which cannot be proved but are seen almost every time.

For example, let me give you a small example. So if you observe the price of any stock, then we often say that the price increases and at some point, it should slightly reduce and then increase again. So the price need not go on increasing at a stretch. There always have to be dips that come in. So all these are stylized facts.

So stylized facts are such facts that cannot be proved. But they are visually observed or seen evidently almost every time. Firstly, thick tails. So, thick tails means that the tails, if you observe, if you have changing variance problems, then as discussed before, one cannot apply a normal distribution to the errors, and hence the tails are thick. They tend to be leptokurtic.

By the way, these are stylized facts of asset returns. So, asset returns have thick tails. Second is the leverage effect. So, the tendency for changes in stock prices to be negatively correlated with changes in volatility. So, what this means is that changes in the actual stock price are negatively correlated with changes in volatility.

So, if you have a highly volatile kind of time point, then changes in stock price would be kind of high again, okay? The third one is the non-trading period effect. So, whenever a market is closed, information seems to accumulate at a different rate than when it is open. Obviously, you might see that markets could be highly volatile on a Monday rather than on a Tuesday or Wednesday when the markets have been open that week. Because all the information sort of accumulates over the weekend.

So, the point where the market closed on Friday could be completely different from the point where it opens on Monday. So, one can actually observe slightly more volatile periods on Monday compared to, let's say, Tuesday, Wednesday, or Thursday, etc., okay? And lastly, forecastable events. So, volatility is high at regular times, such as news announcements or other expected events, or even at certain times of the day. Let's say, for example, less volatile in the early afternoon.

So again just to give you a small example if you I am very sure that if many of you are sort of interested in let's say investing or trading something like that. So, you should have observed that whenever the market opens for the day right. So, the underlying price or the underlying asset or the underlying price of the index is highly volatile as compared to somewhere down the line let us say 12 o clock or 12.30 in the afternoon ok. So, volatility is high at regular time such as whenever let us say news announcement. So, whenever some erroneous news has to be expected the markets are highly volatile.

or rather other expected events or even at certain times of day. So as discussed, let's say highly volatile in the morning whenever the market bell occurs and slightly lesser volatile in the early afternoon. And again, towards the end, whenever the market closes, again, you might see some volatile aspects. So, volatility again is sort of not constant over the day.

So, it might be high in the morning, then sort of subsides down in the afternoon, then again sort of picks up in the afternoon, late afternoon, etc. Now, the next one is volatility and serial correlation. So, there is a suggestion of an inverse relationship between the two. So, if you talk about volatility and serial correlation, then one assumption is that there has to be some inverse or negative relationship between the two. And lastly, co-movements in volatility.

So, there is considerable evidence that volatility is positively correlated across assets in a market and even across markets. So, let us see if you pick up two assets in a same market. then there is considerable evidence that volatility is positively correlated. So I will give you a simple example. So think of these two coins.

So on one hand you have Bitcoin, on the other hand you have Ethereum. And whenever there is high volatile structure in Bitcoin, you should see the same kind of structure in Ethereum and vice versa. So, since these two asset classes come from the same cryptocurrency market and then whenever you see some highly volatile time points in Bitcoin, you will definitely see some same kind of a structure or highly volatile time points in the Ethereum in the same duration. So, this is basically called as co-movements in volatility. So, two assets kind of co-move in terms of their volatile nature or volatility.

Okay, now the next thing is a small primer on types of volatility. So how many kinds of volatility are there, right? So we will sort of describe that for a moment and then proceed ahead. So in fact, you have four different kinds of volatility. So the first one is called as historical volatility.

So, firstly, what do you mean by that? So, what do you mean by historical volatility? So, calculated using the past data, it could be either returns or prices, right? So, the past data could be either returns or prices, right? The second point is, historical volatility represents the standard deviation of returns over a specific time window in the past.

And again, I have probably—if not in this lecture, maybe in the next lecture—I will sort of elaborate a bit more on historical volatility. So, how do you find it out? What are some of the limitations of that? But again, just to understand right now, historical volatility is nothing but a measure of the variance that has occurred in the past. So, calculated using the past data or the past returns or prices.

The second kind of volatility is what is called implied volatility. So, what do you mean by that? So, implied volatility is derived from option pricing models. For example, Black-

Scholes. So, many of you might know that Black-Scholes is a really famous kind of option pricing technique, right?

So, such implied volatilities are kind of derived from any of the option pricing models. And what it does is that implied volatility represents the market's expectations of future volatility. So rather than taking or finding variance from the past data, what do you expect the volatile nature in the market to be in the future? So the market's expectations to capture the future volatility are given by implied volatility. The third one is called volatility clustering, as discussed even before, that financial time series often exhibit periods of high volatility followed by periods of low volatility.

So this phenomenon is called volatility clustering. And this is a very common characteristic in asset prices and is captured by some models that we will study later on, like ARCH and GARCH, so more on them later. So, these two are specific time series models which try to capture such volatility clusters. And the last kind of volatility is called realized volatility. So, what do you mean by realized volatility?

So, the actual volatility observed over a past period. So, realized volatility is nothing but the actual volatility observed over a past period in history. And such volatility can be estimated using any high-frequency data. So, let us say minute-by-minute or daily returns, etc. Does that make sense?

So, again, just to summarize very quickly. So, let me go back a slide. And then the first kind of volatility is called historical. The second one is implied. The third one is volatility clustering.

And then the fourth one is called realized volatility. So, again, each of these different kinds of volatilities has its own advantages, its own ways of finding volatilities, and its own pros and cons when it comes to capturing volatility. For example, again, if you go back a slide, the historical volatility revolves or rather focuses entirely on the variance of the past data, or rather the variance of the past returns or prices, etc. Historical volatility represents the standard deviation of returns over a specific time window in the past. Implied volatility—the idea is entirely different. Such volatility is derived from, let us say, some option pricing models, for example, Black-Scholes or probably some other option pricing model, etc.

And such volatility represents the market's expectations of any future volatilities. And thirdly, again, just to summarize, volatility clustering. So, this is kind of the pain point here. So, this is exactly where we will try to put forward some time series models. For example, ARCH, GARCH, because such volatility clustering is often seen in, let us say, financial time series or day-to-day financial markets, etc.

By the way, just to tell you the full form of ARCH. So, ARCH is autoregressive conditional heteroscedasticity. So, autoregressive conditional heteroscedasticity and G-ARCH or GARCH is nothing but generalized ARCH. So, there has to be some autoregressive component to capture what? To capture the changing variance.

And what kind of variance? Conditional variance. So, auto-eliginous conditional heteroscedasticity. So, heteroscedasticity means that there should be some changing variance aspect and what kind of variance conditional variance and how are you modeling that using some AR model ok. So, one can either use the AR model or the GAR model to model such changing variance aspects and again the last one is realized volatility.

which is kind of similar to let us say historical volatility, but this is the actual volatility observed over a past period and can be rather estimated using some high frequency data. Okay, so now probably in the next lecture, we will talk, we will try to elaborate a bit more on historical volatility. We will try to put forward some limitations of that. And down the line, we will try to elaborate a bit more about one arch model, which is let us say arch 1. And then down the line again, we will discuss a few extensions of arch, a few more extensions of garch, etc.

Thank you.