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 49: GARCH Model Extensions

Hello all, welcome to this course on time series modeling and forecasting using R, ok. Now, again, this is the last session for this week—obviously, the theoretical session—and of course, in the next session, it will be a practical session in R. So, again, just to quickly summarize where we stand before we discuss anything new in this session. So, this entire week has been devoted to exploring how you model the changing variance of a time series or, rather, how you apply some volatility modeling, ok. And again, volatility means changing variance.

So, let us say, again, just to give you a very simple example: if you assume a stock price. And the stock price behaves like this initially, but due to some unforeseen news, you see lots of fluctuations in between. And again, you have a lot of activity down the line. So, we will say that if you observe the overall variance of this plotted data or, rather, plotted return series, then you can see some changing variance in that. So, rather than only focusing on describing the mean of the model—which we have been doing so far using ARMA models, ARIMA models, SARIMA models, etc.—but rather than only doing that, can you also try to model the volatility in the underlying time series or can you try to model the changing variance of the underlying time series, ok?

And in this period, we have been exploring or talking about different sorts of models. The first model we talked about a couple of sessions back was the ARCH model. Which stands for autoregressive conditional heteroscedasticity, and heteroscedasticity stands for changing variance. And just in the last session, we extended the ARCH model because, again, we have seen that the ARCH model may not be a very wholesome sort of model in different situations. So, the extension was coined as GARCH, and the full form of GARCH is generalized ARCH or generalized autoregressive conditional heteroscedasticity.

And again, we saw some limitations of GARCH in the last session—some key features of the GARCH model in the last session. Then, how do you fix the optimal orders, right? So, all these were a few things we have seen in the last session, and right towards the end or somewhere in the middle of the last session, I mentioned that in the next session, which is this one. We will talk about some other extensions of the GARCH model as well because what happens— So, a simple GARCH model may not be able to capture the asymmetry. So, if you remember—I am not sure how many of you remember—again, if you do not remember, you might have to go and rewatch the previous sessions. But again, the GARCH model, or rather let us say a simple ARCH or GARCH model, is not able to capture the asymmetries in the underlying practical data.

Now, firstly, what do you mean by asymmetries? Asymmetries mean that if you have a negative shock or some negative news. So, negative news has a much more drastic impact than positive news. And of course, this is very widely seen in, let us say, stock markets as well.

So, if a company declares some positive news, then the market may not be, let us say, up by a whole lot. So, the market may move up by, let us say, 5 percent, 10 percent at the max. But if some negative shock comes or some negative news comes pertaining to some company, then Then again, we have very well seen in the past that the price has a tendency of dropping, let us say, quite a lot. So, let us say 10 percent, 15 percent, 20 percent, right? Which means that negative news or rather a negative shock can create much more turbulence than some positive event or positive news, okay?

So, for all such ideas, a simple ARCH model or a simple GARCH model may not do the job. So, we have to study some extensions of the standard GARCH model, okay? So, we will study all these extensions one by one in detail—do not worry—but first, a brief background as to why we have to study the extensions again. So, specific variants of the GARCH model exist in literature, tailored to capture different characteristics in the underlying time series data. So, each and every different practical time series data is not the same, of course.

So, how do you tailor to capture the different characteristics in all these different sorts of time series data? Of course, one cannot put forward a simple GARCH model for each and every one of those. So, one would require some specific variants of the GARCH model as well. Particularly, volatility dynamics in financial and economic contexts. So, a few examples are these.

So, GJR GARCH, and then we will talk about exponential GARCH, then we will talk about threshold GARCH, then the next one we will talk about is asymmetric power GARCH, then fractionally integrated GARCH, multivariate GARCH, and lastly, non-linear GARCH. And, of course, If you look through the literature, one can actually see a lot of other extensions as well, but all those extensions kind of stem from one of these that I have listed here. So, predominantly, these are the important ones. So, GJR GARCH, E GARCH, threshold GARCH, asymmetric power GARCH, etc.

So, during the course of this entire session over the next few minutes, we will talk about each and every extension listed here in a bit more detail, along with some examples corresponding to each of the extensions. So, now the very first one is GJR GARCH. So, GJR GARCH—the full form is Glosten, Jagannathan, and Runkle GARCH. So, by the way, GJR stands for the acronym of these three people. So, Glosten, Jagannathan, and Runkle, who put forward this model.

So, in short, GJR GARCH. So, what is the property here? So, GJR GARCH captures the asymmetric effects in volatility. So, as discussed before, a simple GARCH model fails to capture the asymmetries, but GJR GARCH captures the asymmetric effects in the underlying changing variance—where, just to summarize, negative shocks or bad news affect volatility differently than positive shocks or good news. So, how can you capture the asymmetry?

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \gamma\epsilon_{t-1}^2 I(\epsilon_{t-1} < 0) + \beta\sigma_{t-1}^2$$

So, using this equation. So, the model definition is as follows. So, sigma t square equals omega—omega is the same as before. So, omega is the intercept or the constant term, plus alpha multiplied by epsilon t minus 1 square. So, epsilon t minus 1 square is the squared residual, plus gamma multiplied by epsilon t minus 1 square—but this is the extra bit here.

So, what we are doing is we are putting an indicator function. So, again I am not sure how many of you are familiar with the indicator function notation. So, again indicator function is not a very difficult thing to sort of analyze. So, I will give you a small example which is outside of time series. So, let us say if I write down capital I of a random variable x being let us say less than 0 ok.

So, let us say x is a random variable. So, for a second come out of a time series literature. So, this is a very general sort of a random variable in a stats literature. So, if I have this particular indicator function that capital I x is less than 0. So, this would take two values.

So, it would take one value of 1 whenever this event happens which means whenever x is less than 0 otherwise it will take a value of 0. So, any indicator function can only take two values either 1 or 0 and then when would it take a value of 1 whenever this underlying event happens and otherwise it will take a value of 0. Similarly, coming back to this equation, so now, if you want you can pause the video and try to answer the question as to when would this indicator function take a value 1 and when would it take a value 0, ok. So, simply put, so this indicator function would take a value 1 whenever this residual epsilon t minus 1 is less than 0 and otherwise will take a value 0. So, I will say that this is the extra part which was not seen in the Garch model right.

So, again, omega plus alpha epsilon t minus 1 square plus gamma epsilon t minus 1 square, but you have some restriction here that the residual should be negative. Plus beta into sigma t minus 1 square, ok. So, again, one important point is that if you observe the order in the model. So, the order in the model is nothing but 1, 1. Can you see that? Because you do not have summations anywhere, ok. So, just as a simplicity, I have tried to ignore the summation signs and then I fixed the orders to be 1, 1. But of course, if you want to extend this further to multiple orders, I can involve, let us say, something like alpha i or gamma i's or beta i's, etcetera, but of course, there will be summations there.

Hopefully, this point is clear that rather than making the overall equation difficult-looking, I have restricted myself to fix the orders to be 1, 1 wherever possible. So, again, just to summarize quickly, this is the GJR GARCH model, and this is exactly what we talked about a short while back—that the indicator function epsilon t minus 1 being negative is an indicator function equal to 1 if the previous residual is negative and 0 otherwise. And what exactly is gamma? So, gamma, which is assumed to be positive, captures the impact of negative shocks. So, if the past residual is negative, which would correspond to a negative shock, this is the coefficient that captures the impact of that negative shock.

Now, lastly, what exactly could be an application? So, financial markets where bad news tends to increase the volatility more than good news—for example, the leverage effect, ok. Now, again, this is a very, very, very similar kind of application we have been seeing throughout this week, right. So, the leverage effect means that whenever bad news tends to increase the changing variance more than good news, ok. Alright.

So, the first model is GJR-GARCH. Now, coming to the second model. The second model is called E-GARCH, OK. So, E-GARCH, or rather the exponential GARCH.

So, E-GARCH means exponential GARCH. So, E-GARCH handles the leverage effects again. So, again, if you see briefly, both in GJR-GARCH or E-GARCH or some of the other extensions studied today. So, the common idea is how each of the extensions tends to handle the asymmetries. So, again, E-GARCH handles the leverage effects and ensures that the conditional variance is always positive without requiring non-negativity constraints on the parameters.

$$\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \alpha \frac{\epsilon_{t-1}}{\sigma_{t-1}} + \gamma \left(\left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| - E\left[\left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| \right] \right)$$

So, there is one thing extra here: E-GARCH assumes that the conditional variance is always positive. without requiring any non-negativity constraints on the parameters, OK. And this is the model definition. So, log of sigma t squared. So, rather than simply putting sigma t squared, I am taking the log of sigma t squared again equal to some slightly advanced-looking equation.

So, it starts with omega again. So, omega is the constant plus beta multiplied by the log of sigma t minus 1 squared. Again, this is a small change here compared to the GARCH model or GJR-GARCH, etc., as we are taking the log of the past lag of the conditional variance as well, plus alpha multiplied by this ratio. So, this ratio is completely new.

So, the ratio of what? So, the ratio of epsilon t minus 1 divided by sigma t minus 1, and then plus the last coefficient, gamma, multiplied by again the absolute value of epsilon t minus 1 by sigma t minus 1 minus its expected value. So, again, I will not spend time describing the nature of the equation because the nature of the equation is not very easy to understand, but again, the whole point is to note a couple of important points here. So, firstly, again, EGARCH is capable of handling asymmetries or leverage effects, and the second point is EGARCH ensures that the conditional variance is always positive. Which is an advantage, of course, right?

Because otherwise, we would have required some non-negativity constraints on the parameters, stuff like that, okay? But EGARCH does not require any of those things because it ensures that the conditional variance is always positive. And how is it doing that? By taking logs. So, the logarithmic transformation ensures the positivity of the conditional variance term, which is sigma t squared, while gamma, assumed to be positive, captures the impact of negative shocks as before.

So, this gamma stays the same. So, again, gamma has a tendency of capturing the impact of the negative shocks, while since you are taking a logarithmic transformation of sigma t square, it ensures the positivity of the conditional variance term, which is sigma t square. And lastly, the application part. So, again, financial assets with pronounced leverage effects, such as equity indices, etc. So, rather, I can actually try to put forward an EGARCH model or a GJR GARCH model, and then down the line, we can compare.

So, which model suits the underlying data better. Now, the third one is called the threshold GARCH. So, again, a small disclaimer: the short form is T-GARCH, the full form is threshold GARCH, and T-GARCH is similar to GJR GARCH, differing only as it involves a standard deviation term in place of the variance. So, we will see exactly how. So, this is the model definition again, just to summarize that T-GARCH is not different widely from GJR GARCH. The only difference is in GJR GARCH, we are exploring epsilon t minus 1 square, while when it comes to the threshold GARCH, we are taking the standard deviation term in terms of epsilon t minus 1 without the square. So, we will see what I am saying. So, this is the model definition. Again, how do you write down the equation?

$$\sigma_t^2 = \omega + \alpha|\epsilon_{t-1}| + \gamma\epsilon_{t-1}I(\epsilon_{t-1} < 0) + \beta\sigma_{t-1}^2$$

So, on the left-hand side, we are trying to model the conditional variance as always, which is sigma t squared, and then sigma t squared equals omega plus alpha multiplied by the absolute value of epsilon t minus 1 plus gamma multiplied by epsilon t minus 1 indicator function. Again, the same thing. So, epsilon t minus 1 is negative plus beta multiplied by sigma t minus 1 squared. So, this is the model definition. Again, just to summarize very quickly, instead of epsilon t minus 1 squared, let us say here or rather there, we are taking the standard deviation.

So, this is the only difference between GJR-GARCH and PGARCH. Again, when it comes to application, from a financial markets point of view, financial markets where extreme negative shocks disproportionately affect the volatility. So, as you see, broadly speaking, the T-GARCH, E-GARCH, or G-GARCH revolve again around the idea of capturing the asymmetry, which is also called the leverage effect, etc. Now, the next one is called AP-ARCH. So, AP-ARCH, the full form is asymmetric power ARCH.

So, AP-ARCH or asymmetric power ARCH captures asymmetry and allows for flexible modeling of power transformations of the volatility. So, exactly how? So, this is the model definition. So, I guess the name itself speaks as to where the power would be, right? So, again, asymmetric power ARCH.

$$\sigma_t^{\delta} = \omega + \alpha(|\epsilon_{t-1}| - \gamma\epsilon_{t-1})^{\delta} + \beta\sigma_{t-1}^{\delta}$$

So, initially rather than taking sigma t square, we will upfront take sigma t to some power delta. So, sigma t to the power delta again equals the constant term omega plus epsilon plus T minus 1 absolute value of that minus gamma into epsilon T minus 1 to the power delta attached to the coefficient alpha, ok. So, again all the coefficients are as we saw earlier as well. So, alpha, beta, gamma, etcetera—the only difference is how you put forward the

lagged residual term and the lagged conditional variance term. So, whether you take variance, whether you take standard deviation, whether you raise it to some power, or whether you take absolute value. So, these are some small differences that all these extensions have, right? For example, again just to summarize, sigma t to some power delta equals omega plus alpha into absolute value of the standard deviation, which is epsilon t minus 1 minus gamma into epsilon t minus 1 to the power of delta. plus beta into sigma t minus 1 to the power of delta.

where delta is nothing but the power parameter which determines the transformation of conditional volatility, while gamma is the asymmetry parameter. So, gamma is the asymmetry parameter, while delta captures the power parameter. Again, what is the application? So, markets with non-linear and asymmetric volatility patterns, such as energy or commodity markets. So, these are some of the applications where an APR model seems suitable.

Now, the next one is slightly different and is called fractionally integrated GARCH or, in short, FIGARCH. So, FIGARCH or fractionally integrated GARCH, and the idea here is slightly different, as discussed a short while back, that FIGARCH tries to model the long memory property in the underlying volatility. So, rather than capturing the asymmetry, it tries to capture a slightly different thing, which is long memory. And again, long memory, as discussed before, almost all of you must know what we mean by long memory. So, long memory means that if a time series reverts back to its mean very, very slowly in the long term, then we will say that time series has some persistence down the line or it captures some long memory properties.

So, again, in short, this FIGARCH model captures the long memory in the volatility, capturing persistence which decays very, very slowly over time. And this is exactly how it is capable of doing it. So, the model definition combines fractional integration with GARCH dynamics. So, it is a combination of two things. So, fractional integration and some GARCH dynamics, okay.

$$\phi(L)(1-L)^d\sigma_t^2 = \omega + (1-\beta(L))\epsilon_t^2$$

So, again, if you vaguely remember, right. So, again, if you do, then it is a very good thing for me, at least, that again, if you vaguely remember the idea of how we wrote down the fractionally integrated models. And this is exactly one way of writing it down. If you look at the left-hand side of this equation. So, there has to be some coefficient.

So, $\varphi(L)(1 - L)^d$ applied on σ_t^2. So, this is the operator which is applied on σ_t^2. Of course, $\varphi(L)$ is the set of coefficients, right. And $(1 - L)^d$ is the operator applied on the conditional variance, equal to sort of a similar thing. So, ω + (1 − set of coefficients $\beta(L))\varepsilon_t^2$.

Where this $(1 - L)^d$ is a fractional differencing operator with d assumed to be strictly between 0 and 1. So, again, just to summarize quickly, this FIGARCH is nothing but a combination of some fractional integration with the underlying GARCH dynamics. And again, when it comes to application point of view, one can see applications in financial markets where long-term persistence in volatility exists, such as bond yields or exchange rates, etc. So, if you want to account for two things—long memory or persistence along with trying to model the GARCH dynamics—then probably FIGARCH would be much more suitable. Alright, now the next one is M-GARCH.

So, M-GARCH stands for multivariate GARCH, right. So, M-GARCH, as the name suggests, is called multivariate GARCH, which extends GARCH to multiple time series, capturing the co-movements in the volatilities, right. So, let us say if you have more than one time series to capture, right. So, let us say if you have two time series to capture or three time series to capture. So, in that case, a simple GARCH model will not do the job, right, because the GARCH model is univariate.

So, you have to shift attention to some bivariate sort of model or multivariate sort of model. And then, in this regard, the M-GARCH model or multivariate ARCH model extends the GARCH to multiple time series, capturing the co-movements in volatility. And these are some of the popular specifications or some popular models under the M-

GARCH framework. So, the first one is the BEKK model. So, what it does is directly model the variance-covariance matrix, but it suffers from over-parameterization.

So, by the way, these three are models under the M-GARCH model. So, again, we will not go into detail. So, I just thought of putting a slight variant of univariate GARCH, which is called M-GARCH. So, the second model could be the BEKK model. So, it imposes structure on the variance-covariance matrix for more parsimonious modeling.

And lastly, the DCC GARCH or dynamic conditional correlation GARCH. So, it models time-varying correlations alongside volatilities. So, the last one could be seen very often when it comes to, let us say, the actual application point of view. So, the DCC GARCH, called the dynamic conditional correlation GARCH, and what it does is it sort of models two things. So, the time-varying correlations as well as the volatilities.

And lastly, N-GARCH. So, N-GARCH stands for non-linear GARCH, which introduces non-linearity to capture more complex dynamics in volatility. And this is exactly what the model definition looks like. So, it combines the fractional integration with GARCH dynamics, right? So, again, let us say sigma t squared equals omega plus alpha epsilon t minus 1 squared plus beta sigma t minus 1 squared plus lambda epsilon t minus 1 into sigma t minus 1.

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \lambda\epsilon_{t-1}\sigma_{t-1}$$

So, I guess there is a small typo here. So, you should erase this line because it has nothing to do with fractional integration. So, completely ignore that line. So, the model definition is correct, but the description is not correct. So, non-linear GARCH—the whole idea is how you capture the non-linearity in the underlying dynamics of volatility.

And here, can you see how this is capable of capturing the non-linearity by this multiplication term? So, what you are doing is multiplying both things. So, past residual as well as the past conditional variance to involve some complex dynamics in changing variance, and this is the attached parameter to do that. So, lambda is a parameter to capture the non-linear interactions. So, we are having some interaction between the residual portion and the variance portion.

And from the application point of view, so high-frequency financial data with complex volatility behavior. So, this is one application of NCARCH. And then we could have one more extension called HRARCH. So, HRARCH is called heterogeneous ARCH, which explains volatility using returns aggregated over multiple time horizons. So, again, a slightly different kind of idea.

$$\sigma_t^2 = \omega + \sum_{k=1}^{K} \alpha_k \left(\frac{1}{k} \sum_{i=1}^{k} \epsilon_{t-i}^2 \right)$$

So, this is the model definition again. You can ignore this line; it has nothing to do with fraction integration, but this is the model definition. So, sigma t square equals omega. Plus summation k going from 1 to capital K alpha k into 1 by k summation i going from 1 to k epsilon t minus 1 square. So, the whole idea behind H-ARCH is that H-ARCH is capable of explaining volatility using returns aggregated over multiple time horizons. So, it captures the idea that market participants operate over heterogeneous time scales, OK?

So, let us say if you do not have some similar time scales, we have heterogeneous time scales, then what sort of model is suitable in capturing that idea, right? So, one example is HRH, and again, the application of that would be seen in high-frequency trading and markets with participants reacting to news over different horizons. And then this last slide sort of tells you the summary of all the applications. So, let us say GJR GARCH or T-GARCH, then the application could be equity markets where the leverage effect is seen, right? Then exponential GARCH, let us say stock indices or currencies with strong asymmetry in volatility response, right?

Or APR, let us say commodity and energy markets with flexible volatility dynamics. Or FIGARCH, where you can see some long memory in bond yields or currency volatility, etc. And lastly, MGARCH, which would be seen in applications in portfolio risk management or studying inter-asset relationships. So, this is the only one which is different in the sense that it extends from one time series to multiple time series. And hence the name multiple time series.

Or rather, multivariate GARCH or, in short, M-GARCH. So, I try to put this, which is slightly different because it is termed as E-GARCH X or GARCH X, which extends GARCH by incorporating some extra variables in that, and these extra variables are called exogenous variables, which are Xt and how exactly. So, let us say Xt is some extra independent variable. Okay. So, let us say you create some time series; you have some univariate time series running along, and on top of that, you have some extra variable which sort of affects the time series as well.

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \, \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \, \sigma_{t-j}^2 X_t$$

So, one can use the EGARCH X model for that. Again, you can erase this line. So, this is the model definition. So, sigma t square equals omega plus the same summation in terms of epsilon t minus i square, and this is the extra term which is attached to the conditional variance, which is xt. So, what exactly is xt?

So, xt is the exogenous covariate affecting the changing variance. And then, an application could be modeling volatility with macroeconomic indicators, news sentiments, or financial stress indices. Alright, so lastly, we will discuss one very important extension—another important extension—which is called the ARMA plus GARCH model. So, what one can have is a combination of ARMA plus a GARCH model. So, an ARMA plus GARCH model combines the strength of an autoregressive moving average (or, in short, ARMA) model for modeling the mean of a time series with a GARCH model for modeling its conditional variance or volatility.

This hybrid approach captures two things. So, serial correlation via the mean model (which is ARMA) and time-varying volatility via the GARCH structure. So, this is a combination of two things. So, ARMA plus GARCH. And how exactly?

So, again, both equations would be pretty straightforward. So, this is exactly how the ARMA equation runs. This is nothing but the usual ARMA (P, Q) model, right? So, yt equals mu plus the AR part plus the MA part plus the random error. And when it comes to capturing the variance, we can capture the variance using a GARCH (R, S) model, right, using this structure: sigma t squared equals omega plus the ARCH term plus the GARCH term.

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \, \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \, \sigma_{t-j}^2$$

Make sense? So, this is a two-level model. So, you are trying to capture the mean using ARMA, extracting the residuals from the ARMA equation, and then trying to incorporate those residuals in the changing variance structure. And lastly, we have this last slide capturing some applications. So, the first application would be seen in finance.

Let's say modeling stock returns with serial correlation and volatility clustering. Second, it could be seen again in finance, which is forecasting risk metrics such as, let's say, VAR or value at risk. Or one can see it in macroeconomics, let's say modeling GDP growth or inflation with time-varying uncertainty. Or in energy markets, where let's say one focuses on capturing volatility in commodity prices like oil or electricity. So, all these could be applications of RMR plus GARCH sort of a setting.

So, with this, we come to the end of the entire umbrella of modeling volatility and modeling changing variance using different sorts of extensions of GARCH, ARCH, E-GARCH, GJR-GARCH, et cetera, right? So, I think this particular week, we sort of witnessed lots and lots of models, right? So, all the extensions of GARCH—we started with ARCH, then the actual GARCH, then GARCH-1-1 or GARCH-P-Q, and all the extensions we've seen in this lecture. Now, in the next lecture, we will pay attention to a practical example where we will tie all the things down by using some simulated data and practical data.

Thank you.