GARCH and GARCH derived models for volatility modeling and forecasting

Tommaso Grandi

January 2024

Contents

1 Abstract

In finance, volatility is a pivotal metric, serving as a crucial indicator of the degree of variation in asset prices. Essential for effective risk management and asset allocation, it mirrors market uncertainty and significantly influences investment strategies, decision-making, and the pricing of financial products. This essay presents a synthesis of the work of those pioneers who developed auto-regressive time series techniques for analyzing and forecasting market and financial assets' volatility. First, an introduction will highlight the main issues that led to the development of ARCH and GARCH, justifying the need for these models in modern financial modeling. Secondly, we will present the renowned GARCH model that followed the ARCH model developed at the beginning of the eighties. These two models formalized the analysis and modeling of the heteroskedastic nature of financial returns; by describing the variance of the innovations as a function of the previous forecasted variances (only for GARCH), in addition to the magnitude of its past lagged error terms (ARCH). Following on, the discussion will move towards the family of GARCH models: extensions of GARCH that were developed in the following decades and that extended the capability of volatility models in dealing with market volatility asymmetries; anomalies in the distribution of the returns that are typical in periods of financial distress or future uncertainty. The third section will hence present asymmetric models such as the exponential GARCH

(EGARCH) and the GJR-GARCH. Empirical evidence from the literature will then highlight these models' performance, advantages, and disadvantages, and conclusions will be drawn by comparing the discussed models with more sophisticated and modern techniques that feature Deep Learning architectures.

2 Introduction

Financial time series present many anomalies in their distribution of returns, especially when compared to a standard normal distribution. Statistical regression models on the other hand, often consider implausible assumptions that in the past were reasonably not discarded due to the lack of techniques that could allow to correctly model the anomalies. As we may recall, the OLS model assumes that the expected value of squared errors remains constant at any given point in time, when considering a time series. This assumption is called homoskedasticity. However, in the presence of significative heteroskedasticity, wrong estimates for the standard errors and confidence intervals lead to bias in the evaluation of the model, giving a ”false sense of precision”, as Engle (2001) [1] note. This problem consistently characterizes financial time series. Figure 1 shows the scatter plot of the returns of the FTSE MIB 100 index for the Italian stock market. Is sufficient in this case, as well as with other financial time series, to visually inspect the chart to notice that the behavior in the squared deviation of the returns from its long-run mean (i.e. the variance) changes over time, suggesting that some periods feature higher risk than others, and those with higher (or lower) risk are not randomly scattered but are auto-correlated at some degree and even significantly clustered.

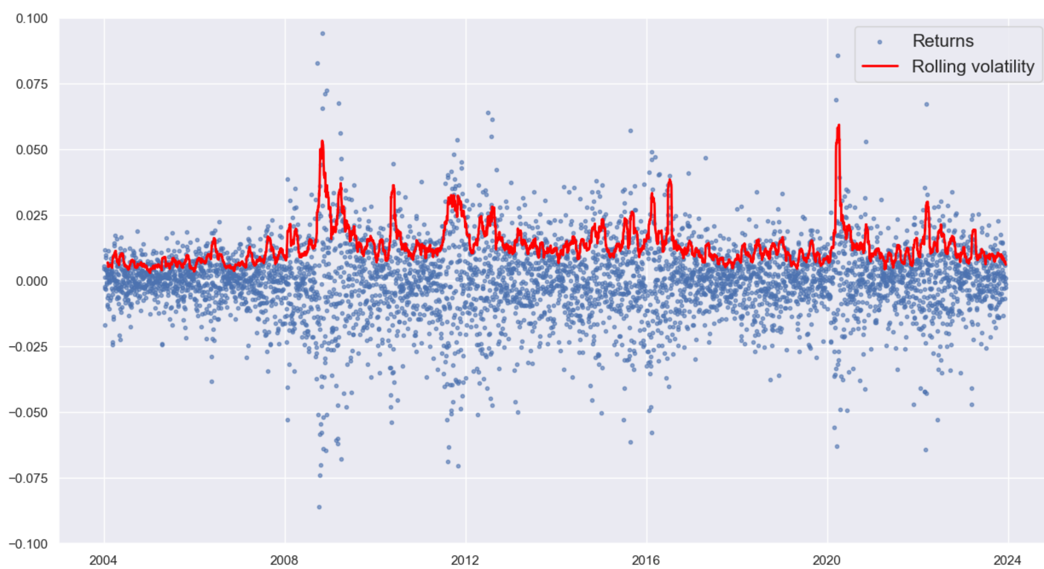

Figure 1: FTSE MIB 100 returns and monthly rolling volatility (21 days window), 2004-2024

In the context of financial modeling and risk management, volatility is commonly considered a measure of the variance of asset prices' returns and can quantify the risk related to a particular financial asset. It's obvious however that this risk measure becomes flawed when taking into consideration the remark about heteroskedasticity. Additionally, we highlighted how the series in Figure 1, as well other financial series, are characterized by volatility clustering, that is, extended periods of 'violent' or high market volatility followed by a period of high volatility, and 'calm' or low market volatility followed by a period of low volatility (Tsay (2010) [2]).

As a result of this "volatility clustering", models capable of handling time-varying volatility and volatility clusters became necessary. Engle (1982) [3] was the first to develop an autoregressive model that could capture this kind of anomalies in the volatility of financial time series. In an effort to model the inflation of United Kingdom without assuming constant volatility over time, he proposed the ARCH model, which allows an auto-regressive (AR) component of order p to enter a variance process describing the volatility of returns' innovations. In the paper, the structure of the model is defined as follows. First, a series of random variables y_t conditioned on the information set available at the previous time step Ψ_{t-1} is described as:

$$y_t = \mu_t + \epsilon_t \quad with \;\; y_t|\Psi_{t-1} \sim N(\mu, \sigma_t^2) \tag{1}$$

Intuitively y_t follows some kind of stationary process and the focus of ARCH is to capture its variance, which depends on the variance of the error terms σ_t. Equation 1 usually referred to as *mean process* of the ARCH model. Considering financial assets, y_t usually represents returns obtained by first-order-differencing the original series. Common practice suggests obtaining this series by calculating logarithmic returns. Once the mean process is defined, the included error terms are specified according to equation (2):

$$\epsilon_t = \sigma_t z_t \tag{2}$$

defining the error term as the product of the conditional standard deviation σ_t and a white noise zero mean term z_t with unitary variance. Then, by definition, ϵ_t has mean zero but a conditional variance equals to σ_t, which could be changing through time. The conditional volatility σ_t is what the ARCH model tries to forecast through the autoregressive process of past squared shocks $\epsilon_{t-1}^2, \epsilon_{t-2}^2, ...\epsilon_{t-p}^2$:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + ... + \alpha_p \epsilon_{t-p}^2$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 \tag{3}$$

where p is the farther lagged shock in time (i.e. the order of the ARCH process) and ϵ_t is obtained as $\epsilon_t = y_t - \mu_t$. This *variance process* allows the volatility to be time-varying, and be *conditioned* or dependent on the past squared errors. When past squared errors are larger, the forecasted conditional variance becomes larger. Consequently, the larger conditional variance influences the size of the following errors (equation 2). This means that in the ARCH framework, the variance of the current error is an increasing function of the magnitude of the realized lagged errors. Hence, large shocks tend to be followed by other large shocks, and similarly, small shocks tend to be followed by small shocks, enabling to capture volatility clusters. Of course, we can interpret the size of past errors, as the impact of past news (either positive or negative) on the conditional volatility estimated by the model.

Finally, the estimation procedure is conducted according to the following. First, the errors are obtained from the relation $\epsilon_t = y_t - \mu_t$ by estimating the coefficients of the mean process enclosed in μ (Engle suggests using OLS) and subtract the term from the y_t series. Secondly, these errors are passed into the variance process from equation (3) and the coefficients $\alpha_1, ..., \alpha_p$ are estimated through Maximum Likelihood Estimation. As Engle (1982) [3] explains, MLE is preferred to OLS in this case since "maximum likelihood is different and consequently asymptotically superior;[...] is nonlinear and is more efficient than OLS".

The goal of this model, as explained, is to provide a consistent a reliable measure for the volatility that can be used for issues that concern risk management, portfolio analysis and screening, and the pricing of financial derivatives. But its usefulness resulted not only in capturing volatility issues, but also other important stylized facts of financial data, whose distribution often empirically results in fat tails, negative skewness and leptokurtic shape. This discussion is deferred to the third section, which will showcase different asymmetric models that are designed to be further consistent and adjustable to such anomalies.

It's evident how the structure of ARCH, although simple, became a groundbreaking practice in volatility modeling. Additionally, it has prompted the research in this field becoming the baseline of all autoregressive heteroskedastic models that followed. However, the model's ability of capturing conditional volatility can be limited in some ways. Mainly, empirical research that followed Engle's work showed that "high ARCH order has to be selected in order to catch the dynamic of the conditional variance" (Peters, 2001) [4]. GARCH models became an answer to this issue.

3 Related work

3.1 The GARCH model

Bollerslev (1986) [5] developed and extended Engle's work by conceiving the Generalized Autoregressive Conditional Heteroskedasticity model (GARCH). As introduced before, one of the main issues of ARCH was its greedy nature that requires a high order model to correctly capture the volatility dynamics. This can be an issue when estimating the model's parameters, since having higher-order lags of squared shocks involves estimating many coefficients. GARCH, on the other hand, was developed to be parsimonious by including q lags of the conditional volatility in the variance process from equation 3. Intuitively, lagged forecasts of the conditional volatility, contain the information (news) that in the ARCH model had to be sourced by including a greater number of lags of the squared shocks. Estimation, as it was with the ARCH, is computed with Maximum Likelihood Estimation.

The GARCH(p,q) is represented as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{4}$$

As before, ϵ_t is a 'real value discrete-time stochastic process' (Bollerslev, 1986), the innovations obtained from a regression problem according to $\epsilon_t = y_t - \mu_t$, with μ_t being any form of regression or autoregressive and moving average process (ARMA) estimating y_t. Bollerslev assumed that $\epsilon_t | \psi_{t-1} \sim N(0, \sigma_t^2)$, however, it should be noted that in practice the errors are often assumed to follow other distributions, such as standardized Student-t or a generalized error distribution. This operation can help to capture anomalies in financial time series such as returns. Notably, the tails of the returns distribution are more pronounced, or 'fat', than those of a normal distribution, and hence imply a higher probability of extreme events that are far from the center of the distribution.

Equation 4 involves many conditions in order to be correctly specified, in particular:

$$p > 0, \quad q \geq 0$$

$$w > 0, \quad \alpha_i \geq 0, \quad i = 1, ..., p$$

$$\beta_i \geq 0, \quad i = 1, ..., q$$

For q=0 the process reduce to an ARCH(p). As seen, with GARCH the conditional variance is now not only a linear function of past residuals, but the process allows lagged conditional variances to enter. Calculating the kurtosis of this process, Engle (1982) showed that a GARCH process has higher kurtosity than a standard normal distribution. Consequently, the tail of a GARCH process results in being heavier than that of a normal distribution, as it is frequently observed for financial data.

A very appealing feature of the GARCH concerns the series dependence in ϵ_t. Is sufficient to rearrange equation 4 as:

$$\epsilon_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \epsilon_{t-j}^2 + \sum_{j=1}^{q} \beta_j \upsilon_{t-j} + \upsilon_t \tag{5}$$

with $\upsilon_t = \epsilon_t^2 - \sigma_t^2$, to discover that the GARCH(p,q) process can be interpreted as an autoregressive moving average process in ϵ_t^2 of orders $m = \max(p, q)$ and p, respectively (Bollerslev, 1986) [5]. According to Bollerslev (1992) [6]: "this idea can be used in the identification of the orders p and q, although in most applications $p = q = 1$ is found to suffice". GARCH(1,1) is indeed a highly regarded model by researchers and investment professionals for many empirical applications in the context of volatility modeling. A GARCH(1,1) variance process can be written as:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{6}$$

There is a useful interpretation for the parameters in equation 6. The coefficient ω represents a constant variance that corresponds to the long-run average volatility. The last two coefficients determine how fast the future variance changes with new information and how fast it reverts to its long-run value ω. In particular, the parameter α_1 is the immediate impact of the past shock on the future volatility, whereas β_1 represent the duration of this impact.

From a practical standpoint, GARCH improves the estimation procedure compared to the original ARCH, thanks to the parsimonious specification of the model. In addition, Bollerslev empirically tested a GARCH(1,1) retrieving the problem addressed in the 1983 paper of Engle and Kraft, which concerned the rate of growth in the US' Gross National Product deflator. Testing for the serial correlation in the squared errors, a common practice for checking if residuals are behaving as white noise, resulted in none of the first ten autocorrelation lags exceeding the two asymptotic standard errors, confirming an appreciable behavior of the errors in the model. Additionally, GARCH(1,1) provided a slightly better fit than the ARCH(8) model in Engle and Kraft (1983), and exhibits a more reasonable lag structure (Bollerslev, 1986).

However, as with its former model, GARCH can encounter some weaknesses that have been addressed in its following developments. Primarily, empirical studies of high-frequency financial series have shown that even the leptokurticity captured by the tail behavior of GARCH models remains too weak even when assuming other conditional distributions with fatter tails than the normal, such as the standardized Student-t. Additionally, the main issue that these models are not able to capture is the one anomaly in financial returns first discussed by Black (1976) [7]. ARCH and GARCH models assume that positive and negative news captured by the magnitude of the errors have the same effect on volatility since both positive and negative shocks enter symmetrically in the variance process when being squared. However, according to Black, there is a negative correlation between the present return and the future volatility, a *leverage effect*. This effect implies that there should be some kind of asymmetry in the response of the conditional volatility concerning negative and positive shocks. Empirically, this phenomenon has been observed and confirmed and it has been the main motivation behind the development of the EGARCH model, as well of others.

3.2 EGARCH

The EGARCH model, developed by Nelson (1991) [8], was motivated to address the volatility issues that ARCH and GARCH models were unable to reply. The model represents a successful attempt to account for the excess conditional kurtosis in assets and index returns, as well as the asymmetric response of volatility to positive and negative news. EGARCH exploits the empirical observation of negative correlation between returns and future volatility, first motivated by Black (1976) [7], by allowing the conditional variance to be a function of not only the size of lagged errors but also of the sign. Another limitation of GARCH that the EGARCH tries to address are the non-negativity constraints on the coefficients of GARCH variance process defined in Section 3.1 . Those constraints, as Nelson notes, imply that "increasing ϵ_t^2 in any period increases σ_{t+m}^2 for all $m \geq 1$, ruling out random oscillations in the variance process". Additionally, he argues that non-negativity constraints create difficulties in estimating the model.

The model considers a weighted innovation $g(\epsilon_t)$, that in order to accommodate asymmetry is a function of both magnitude and the sign:

$$g(\epsilon_t) = \theta\epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)] \tag{7}$$

with both ϵ_t and $(|\epsilon_t| - E(|\epsilon_t|))$ being a zero mean i.i.d. random sequences. If $0 < \epsilon_T < \infty$, $g(\epsilon_t)$ is linear in ϵ_t with slope equal to $\theta + \gamma$. Conversely, if $-\infty < \epsilon_t \leq 0$, $g(\epsilon_t)$ is still linear but with slope $\theta - \gamma$. This implies that $g(\epsilon_t)$ allows the conditional variance process:

$$\ln(\sigma_t^2) = \omega + \sum_{k=1}^{p} \beta_k g(\epsilon_{t-k}) + \sum_{k=1}^{q} \alpha_k \ln \sigma_{t-k}^2 \tag{8}$$

to respond asymmetrically to rises and falls in the market or in assets prices. Since $\ln(\sigma_t^2)$ can be negative, there are no sign restrictions on the parameters of the model as there were for a GARCH model, hence EGARCH overcomes the non negativity constraints.

To confirm empirically all the theoretical work behind the EGARCH model, Nelson decided to test it by considering daily returns of the CRSP market index for the period 1962-1987. He employed an AR(1) for modeling the mean process of excess returns, and EGARCH variance process with Generalized Error Distribution, to accommodate for the possibility of non-normality in the conditional distribution of returns. Nelson discovered that the asymmetric relation between returns and changes in volatility, captured by θ, is highly significant and negative, as expected. Additionally, he highlighted how "All the major episodes of high volatility are associated with market drops". The second important observations that he poses regard the observed fatter tails of financial returns. The model generated a significantly thicker-tailed distribution for the errors ϵ_t, thanks to the usage of a GED conditional distribution.

To sum up, Nelson's paper presented an improved ARCH model that didn't suffer some of the drawbacks that characterized the classic ARCH and the GARCH. As Nelson further explains, his goal was to build a model that could allow for the same degree of simplicity and flexibility in representing conditional variance as ARIMA and related models have allowed in representing the conditional means. Thereby, we can confidently say that its goal was achieved.

3.3 GJR-GARCH

Similar to the work accomplished by Nelson; Glost, Jagannathan and Runkle (1993) [9] proposed a model that could account for the asymmetric response of volatility. The three proposed a model, known today as GJR-GARCH, that is just a simple extension of the GARCH model that allows the conditional volatility to have a a separate response to past positive or negative shocks. The variance process in a GJR-GARCH(p,q) is:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p}(\alpha_i + \gamma_i I_{t-i} +)\epsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j \sigma_{t-j}^2 \tag{9}$$

or, for $p = 1, q = 1$, as:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \epsilon_{t-1}^2 I_{t-1} + \beta_1 \sigma_{t-1}^2 \tag{10}$$

where I_{t-1} is an indicator function that takes value:

$$I_{t-1} \begin{cases} 1 & \text{if } \epsilon_{t-1} < 0 \\ 0 & \text{if } \epsilon_{t-1} \geq 0 \end{cases} \tag{11}$$

Thanks to this indicator function the model implements one extra parameter γ_1 into the variance process, which emphasizes the effect of negative past news on the conditional volatility and allows the model to better retrace the asymmetry in the volatility. From equation 10 it's easy to see that positive ϵ_{t-1} contributes with $\alpha_1 \epsilon_{t-1}^2$ to σ_t^2, while negative ϵ_{t-1} have a larger impact of $(\alpha_1 \gamma_1)\epsilon_{t-1}^2$ when $\gamma_1 \neq 0$. Glost, Jagannathan and Runkle motivated the use of this indicator function since "If future variance is not a function only of the squared innovation to current return, then a simple GARCH model is misspecified and any empirical results based on it alone are not reliable". However, even if the model goes in the right direction by trying to account for asymmetry, still takes a step backward from the EGARCH, as its parameters $\alpha_i, \gamma_i, \beta_i$ in equation 9 need to satisfy the same non-negativity constraints of GARCH. Nevertheless, the three researchers empirically tested it in various flavors, some accounting for seasonality components included in a GARCH-M framework (not discussed here), and achieved notable results that underlined a misspecification of GARCH-M and GARCH models; reconciled with the use of GJR-GARCH, which proved a statistically significant negative relation between conditional mean and conditional variance.

3.4 Recent developments in volatility modeling: hybrid GARCH models with Deep Learning

Parallel to the development of the three models just presented, a multitude of derived models were created. Some variations are asymmetric models such as TGARCH and QGARCH; that as EGARCH and GJR-GARCH capture the asymmetric behavior in the response of volatility to positive and negative shocks. Other popular variations are the GARCH-M, that adds a heteroskedasticity term in the mean equation 1; the COGARCH, a continuous time generalization of the GARCH implementing Levy process' increments; and the Zero-Drift GARCH (ZD GARCH) that sets the drift term ω in equation 4 equal to 0.

In recent years, however, the continuous development in the field of machine and deep learning shifted the focus of many researchers toward hybrid techniques that intersect these new methodologies with the field of econometrics. This happened also for volatility modeling, which started benefiting the application of both techniques. This kind of application tries to leverage the positive aspects of GARCH models with the benefits of using deep learning methods, that can capture complex nonlinear relationships that often escape the eye of conventional statistical methods.

One mentionable implementation was the one of Kim and Won (2018) [10], who developed a hybrid model that integrates an LSTM network and multiple GARCH models to forecast the volatility of market indexes. According to the two researchers, the use of an artificial neural network in conjunction with statistical methods offers a clear advantage since it can handle non-linear relationships and is not limited by conventional econometrics assumptions like stationarity. Additionally, the use of feedback connections inside the network, common in Recurrent Neural Networks (RNN), can be effective for sequential data like time series. Furthermore, the architecture that Kim and Won proposed here differs from the previous hybrid applications, combining a single econometric model with single neural network, since it combines the information produced by various statistical models with a RNN model. The statistical models under discussion are the GARCH (powerful for volatility clustering), the Exponentially Weighted Moving Average (suitable for capturing short term changes) and the EGARCH (useful for leverage effect modeling). The procedure for the model is the following. First, the three model parameters are estimated. Once the parameters are estimated on the considered data, is possible to construct a hybrid model by selecting combinations of the three models, and connect their parameters to a LSTM network.

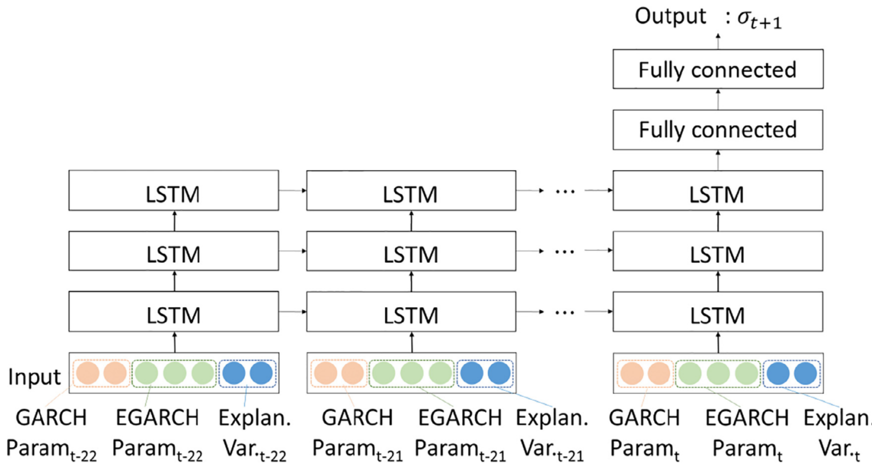

Figure 2: Architecture of a GE-LSTM hybrid model, created by integrating the GARCH, EGARCH, and LSTM models (Kim and Won, 2018 [10])

In particular, if a GARCH-EGARCH-LSTM (GE-LSTM) is considered, as in figure 2, the models' parameters of a GARCH and EGARCH are passed with explanatory variables from the considered period (in this case, the 22 days prior to t) to a fully connected three layer LSTM that predict volatility for the following day $t+1$. The duo implemented and tested various combinations of the model above, in order to assess the superiority of multiple GARCH-type hybrid models (with two or three GARCH variations) with respect to prior hybrid architecture that featured only a single GARCH. Additionally, they tested the combinations of GARCH models with two types of neural network architectures, a basic deep forward network (DFN) and a Long Short Term Memory (LSTM). The best model in terms of predictive accuracy turned out to be GEW-LSTM, integrating all of the three estimated statistical models, a GARCH, EGARCH and EWMA. In general, all hybrid models based on a LSTM network performed better than the single GARCH models and the hybrid models with DFN. The GEW-LSTM improved accuracy in terms of MSE by 76.75 % when being compared with the best performing econometric model, the EGARCH. Furthermore, for LSTM all the multiple hybrid models were better than the single hybrid models. This study clearly shows that the predictive performance of GARCH models, at least for 1 step ahead forecast, can be consistently improved by integrating multiple GARCH-type models with modern deep learning architectures that are designed to capture temporal series patterns. According to Kim and Won the really good performance of the LSTM hybrid architecture is due to the network exceptional ability of learning '*high-level temporal patterns*' when being fed with useful data like the GARCH-type parameters', that contains pieces of information related to volatility; namely the magnitude effect of past volatility on the current prediction and the persistence of the shocks.

Other examples of hybrid models in the literature comes from Koo & Kim (2022) [11] that similarly to Kim & Won, integrated a hybrid GARCH-LSTM that includes a technique named VOLUME-UP, featuring a non-linear filter that transforms the volatility distribution. According to the two, the performance in the prediction of abnormal events of prior hybrid models is limited, being extremely unstable when dealing with financial data characterized by very skewed distributions. As a result, they proposed this novel method that artificially manipulates the volatility distribution and that further enhanced hybrid GARCH models accuracy by 21%.

4 Conclusion and future work

It's clear how Deep Learning hybrid models offer a clear advantage when compared to classical econometric models. However it should be noted that in terms of predictive performance, hybrid GARCH are also overperforming the volatility forecasting accuracy of simple Deep Learning models, as also noted by Kim and Won (2018) [10]. This hence means that, as explained in section 3.4, GARCH models cannot be discarded and have instead a lot of information that can be leveraged by more sophisticated models. Ideally, future endeavors in this field of research should be oriented in finding new hybrid models that can further consistently estimate volatility while being resistent to the acknowledged anomalies that are still hardly solvable with modern architectures. As seen before, hybrid GARCH models are notably unstable when dealing with distributions of financial data that are extremely skewed and hence biased on one side of the distribution. This aspect can cause low prediction accuracy on the opposite side and hence limit the model's adaptability to unfrequent and abnormal events. Possible sources of solution could include the use of Markov Switching Models in a hybrid model scenario. Markov Switching Models have been notoriously applied in finance to identify regime shifts or changes in the underlying structure of the time series, i.e. the underlying statistical distribution. These changes are driven by evolutions in market phases that follow economic cycles or sudden events. Intuitively an architecture that can anticipate these situations and adapt its parameters to be consistent to such rapid variations in the underlying target distribution can improve the modeling and forecasting of volatility. Bildirici and Ersin (2014) [12] have already tested a model implementing Markov Switches in a ARMA-GARCH scenario, however, extensions of this that take into account also hybrid GARCH-Neural Networks models could further improve the predictive accuracy of both solutions and enhance the stableness of the model in case of extreme events that often weaken the solidity of volatility models.

In conclusion, ARCH, GARCH, and GARCH-derived models represent a very important step in the field of volatility modeling, that will be hardly surpassed and totally excluded from future discussions in time series and econometric theory. The models have indeed some weaknesses, but at the same time provide a simple and elegant solution to the complex issues that characterize the heteroskedastic nature of financial data. Hence, practitioners and researchers should embrace the important notions of these models as the starting point for a new generation of hybrid volatility models.

References

[1] Robert Engle. Garch 101: The use of arch/garch models in applied econometrics. *Journal of economic perspectives*, 15(4):157–168, 2001.

[2] R.S. Tsay. *Analysis of Financial Time Series*. CourseSmart. Wiley, 2010.

[3] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

[4] Jean-Philippe Peters. Estimating and forecasting volatility of stock indices using asymmetric garch models and (skewed) student-t densities. 04 2001.

[5] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[6] Tim Bollerslev, Ray Y Chou, and Kenneth F Kroner. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1-2):5–59, 1992.

[7] Black F. Studies of stock price volatility changes. *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, 1976.

[8] Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, pages 347–370, 1991.

[9] Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801, 1993.

[10] Ha Young Kim and Chang Hyun Won. Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models. *Expert Systems with Applications*, 103:25–37, 2018.

[11] E Koo and G Kim. A hybrid prediction model integrating garch models with a distribution manipulation strategy based on lstm networks for stock market volatility. ieee access: Practical innovations, open solutions, 10, 34743–34754, 2022.

[12] Melike Bildirici, Özgür Ersin, et al. Modeling markov switching arma-garch neural networks models and an application to forecasting stock returns. *The Scientific World Journal*, 2014, 2014.