# Modeling the volatility of cryptocurrencies.

Mustapha KHALFOUNI
Doctoral Students
Team applied mathematics
and computer science
Abdelmalek essaadi University
mustapha.khalfouni@etu.uae.ac.ma

Lamarti Sefian MOHAMED
Professor in higher education
Team applied mathematics
and computer science
Abdelmalek essaadi University
lamarti.mohammed.sefian@uae.ac.ma

Noaman LAKCHOUCH
Doctoral Students
Team applied mathematics
and computer science
Abdelmalek essaadi University
statstudy16@gmail.com

***Abstract*—This article examines the intricacies of cryptocurrency financial market volatility, shedding light on the prediction of future volatilities. It explores various statistical models (Multiple Regression, GARCH, EGARCH, etc.) and machine learning (ANN) for forecasting. By utilizing R and/or Python for analysis, the study combines historical data, implied volatility, and external factors such as macroeconomic indicators to enhance the efficiency of predictive models. The aim is to provide investors and policymakers with tools for better risk management and investment decision-making.**

***Index Terms*—Cryptocurrency market volatility, Statistical models in finance, Analysis with R and Python, Historical data and implied volatility.**

## I. INTRODUCTION

Forecasting financial market volatility holds paramount importance for investors, portfolio managers, and policymakers, due to its pivotal role in shaping investment strategies and risk management practices. Volatility, indicative of the range of price movements of an asset within a specific timeframe, is a fundamental metric of market risk and uncertainty. This paper endeavors to investigate the predictive power of historical and implied volatility, alongside additional relevant variables, on the future volatility of assets. Through comprehensive quantitative analysis, this study explores the utility of leveraging data on past volatility and market projections, as encapsulated by implied volatility, for devising potent predictive models. Additionally, it underscores the significance of macroeconomic factors, market indicators, and geopolitical developments as auxiliary elements that affect future volatility. By employing a novel methodological framework that merges sophisticated statistical methods with machine learning techniques, this paper seeks to deepen the understanding of the forces that drive market volatility and to enhance the precision of volatility predictions. This effort aims to furnish stakeholders with crucial insights for optimizing risk management and investment planning.

## II. LITERATURE REVIEW

Predicting the volatility of financial markets is a dynamic field of research that continues to attract the attention of economists, financiers and academics. This literature review focuses on significant contributions in the field, exploring the relationship between historical volatility, implied volatility, and other factors in predicting future asset volatility.

*Historical Volatility as a Predictor*

Historical volatility, calculated from past price fluctuations of an asset, has long been used as an indicator of future volatility. The work of [1] examined the efficiency of volatility forecasts based on historical models, concluding that while these models are relatively simple, they provide a solid foundation for estimating future volatility. However, their performance can be improved by incorporating additional information.

*Implied Volatility as an Alternative*

Implied volatility, derived from option prices, is considered an anticipatory measure of volatility, reflecting market expectations regarding future fluctuations. Studies by [2] and [3] have demonstrated that implied volatility is a more accurate predictor of future volatility than predictions based solely on historical volatility, suggesting that option markets incorporate relevant information not captured by models based on historical data.

*Integration of Additional Factors.*

Beyond historical and implied volatility, other factors have been explored to enhance volatility prediction. The research of [4] on Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models introduced mechanisms for incorporating shocks and market arrival information to predict volatility. Furthermore, recent studies have incorporated macroeconomic variables, market sentiment indicators, and measures of systemic risk to refine volatility forecasts.

*The Impact of Geopolitical and Macroeconomic Events.*

The work of [5] has highlighted the significant impact of oil shocks and monetary policies on financial market volatility, suggesting that geopolitical events and macroeconomic policy decisions are important predictors of volatility.

*Innovative Approaches: Machine Learning and Big Data.*

The advent of machine learning and big data has opened up new avenues for volatility prediction. Recent research, such as that of [6], has explored the use of deep learning techniques to capture complex patterns in market data and improve the accuracy of volatility forecasts.

The literature on volatility prediction highlights a continuous evolution of methods and models, moving from traditional approaches based on historical and implied volatility to more sophisticated models that incorporate a wider range of data and analytical techniques. As research progresses, the goal remains to provide more accurate and reliable tools for risk management and investment decision-making in financial markets.

## III. Prediction models

### A. ARCH Models (Autoregressive Conditional Heteroskedasticity

ARCH models, [7], allow for modeling the volatility of a financial time series by assuming that the current variance is linearly dependent on the squares of past error terms.

### B. GARCH Models (Generalized Autoregressive Conditional Heteroskedasticity)

GARCH models, [8], are an extension of ARCH models that include past volatility terms in addition to the squares of past error terms, allowing for a more flexible modeling of conditional variance. For example, the GARCH (1,1) model can be specified as follows:
*Equation of the mean :*

$$r_t = \mu + \epsilon_t$$

*Equation of the variance :*

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

Where $r_t$ is the return at time t $\mu$ is the mean of return, $\epsilon_t$ is the error term with a mean of 0 and a variance of $\sigma^2$, $\omega$, $\alpha$, and $\beta$ are the parameters of the model to be estimated. The estimation of the parameters is done using the maximum likelihood method, to find the values of the parameters that maximize the likelihood function of the observed data.

### C. EGARCH models (Exponential Generalized Autoregressive Conditional Heteroskedasticity)

The EGARCH [9] models is an extension of the traditional GARCH model, designed to better capture the asymmetries and leverage effects often observed in financial time series data. Unlike GARCH models, which can only capture positive effects on volatility, the EGARCH model allows for both positive and negative shocks to have a differential impact on volatility, thus providing a more nuanced understanding of how news affects market volatility.

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

where:

- $\sigma_t^2$ is the conditional variance (volatility) at time;
- $\omega$, $\alpha_i$ and $\beta_i$ are parameters to be estimated;
- $g\left(\epsilon_{t-1}\right)$ is a function of the lagged error terms $\epsilon_{t-i}$, which typically includes terms to capture the size and sign of the shocks, such as:

$$g(\epsilon_{t-1}) = (\theta\epsilon_{t-i} + \gamma\left(|\epsilon_{t-i}| - E[|\epsilon_{t-i}|]\right))$$

- $\theta$ captures the leverage effect, allowing the model to differentiate between positive and negative shocks;
- $\gamma$ measures the impact of the magnitude of shocks on volatility, irrespective of their sign;
- p and q denote the order of the autoregressive and moving average components, respectively.

### D. Stochastic Volatility Models

These models assume that volatility follows a stochastic process rather than being conditionally heteroskedastic, where conditional volatility is a deterministic function of past information. In SV models, volatility itself follows a stochastic process. This allows for greater flexibility, particularly in capturing sudden changes in volatility and leverage effects. Estimating SV models is more complex than estimating GARCH models due to the stochastic nature of the volatility process. Commonly used estimation methods include. Specification of the Stochastic Volatility Model A basic SV model can be specified as follows:
*Return Equation:*

$$y_t = \mu + \exp\left(\frac{h_t}{2}\right)\epsilon_t, \quad \epsilon_t \sim N(0,1)$$

where $y_t$ is the log-return of the asset at time $t$, $\mu$ is the mean of the log-returns, $h_t$ is the stochastic volatility process, and $\epsilon_t$ is a shock to the returns that is normally distributed.
*Volatility Equation*

$$h_{t+1} = \alpha + \beta h_t + \sigma\eta_t, \quad \eta_t \sim N(0,1)$$

where is $\alpha$ constant term, $\beta$ measures the persistence of volatility, $\sigma$ is the volatility of the volatility shocks, and $\eta_t$ is a shock to the volatility that is also normally distributed and generally assumed to be independent of $\epsilon_t$.

### E. Regime-Switching Models.

also known as "Regime-Switching" models, are a class of econometric models that allow for capturing changes in the regimes or states of an economic or financial process. These models are particularly useful for analyzing data that exhibit structural changes or different behaviors during different periods. They are often applied to the analysis of financial market volatility, economic cycles, and other phenomena where the model parameters can change depending on the current state or regime. *Observation Equation:*

$$y_t = \mu_{S_t} + \phi_{S_t} y_{t-1} + \epsilon_{t,S_t}, \quad \epsilon_{t,S_t} \sim N(0, \sigma_{S_t}^2)$$

Here,$y_t$ represents the observed data at time $t$, $\mu_{S_t}$ is the mean term for the regime $S_t$ at time $t$, $\phi_{S_t}$ is the autoregressive coefficient in regime $S_t$ , and $\epsilon_{t,S_t}$ is the error term for regime $S_t$ , , which is assumed to be normally distributed with mean 0 and variance $\sigma_{S_t}^2$.

*Regime-Switching Mechanism:*

$$P(S_t = j|S_{t-1} = i) = p_{ij}$$

where, $p_{ij}$ represents the probability of switching from regime i at time $t-1$ to regime $j$ at time $t$.

### *F. Machine learning-based models*

Machine learning-based models for volatility prediction are an innovative approach that leverages advanced techniques to capture complex and nonlinear relationships in financial data. These models can incorporate a wide range of features (or explanatory variables) and are particularly useful for modeling financial phenomena where traditional relationships do not apply or are insufficient. Commonly used machine learning models for volatility prediction include:

**Artificial Neural Networks (ANN):** Capable of capturing complex nonlinear relationships through hidden layers and processing units.

**Support Vector Machines (SVM):** Used for predicting financial time series by exploiting high-dimensional feature spaces.

**Random Forests (RF):** A set of decision tree techniques for regression that improves robustness and reduces the risk of overfitting.

**Gradient Boosting:** An ensemble method that sequentially builds a set of decision trees to enhance predictions.

To select the most suitable model, we will undertake an evaluation using the correct metric methods to assess the model's precision. These include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), accuracy, recall, F1 score, or the Area Under the ROC Curve (AUC-ROC) for classification challenges.

## IV. Analysis plan.

To write this analysis article on volatility prediction using the previously mentioned models, we will adopt an analysis strategy organized around the following key steps: Problem definition and objectives:

Specify the objective of volatility prediction (e.g., short or long-term prediction, for specific assets, etc.). Identify constraints and assumptions (e.g., data frequency, considered period). Data collection and preliminary processing:

Choose appropriate data sources (asset prices, trading volumes, macroeconomic indicators, etc.). Clean and organize the data (handling missing data, normalization, etc.). Set the data frequency (daily, weekly, etc.). Exploratory analysis:

Examine the statistical properties of time series (mean, variance, autocorrelation, etc.). Visualize trends, seasonal cycles, and potential change points. Modeling:

Implement the mentioned models to select the most accurate representation of reality. Model validation:

Split the data into training and testing sets. Use performance indicators (e.g., mean squared error, accuracy rate) to assess and compare the models. Model selection and calibration:

Select the most effective model(s) based on validation criteria. Adjust the parameters of the chosen model to optimize its performance. Implementation and practical testing:

Deploy the selected model to make predictions in real-world situations or on new data. Monitor the model's performance and make necessary adjustments. Conclusion and recommendations:

Synthesize the results, discuss the performance of different models. Provide recommendations for the practical application of volatility predictions.

## V. Application

The rapid rise of cryptocurrencies has radically transformed the financial world while creating a new field for data analysis and forecasting. The highly volatile nature of cryptocurrencies, marked by spectacular and unpredictable price fluctuations, presents unprecedented opportunities and challenges for investors, market operators, and financial analysts. In the face of this, the use of advanced predictive models on cryptocurrency data becomes essential for deciphering and predicting market trends. In this context, we will apply these models to the closing prices of three main cryptocurrencies, namely Bitcoin (BTC), Ripple (XRP) and Ethereum (ETH). Before any analysis, the following graphs Fig n°1, Fig n°2 and Fig n°3 clearly show the highly volatile nature of prices.

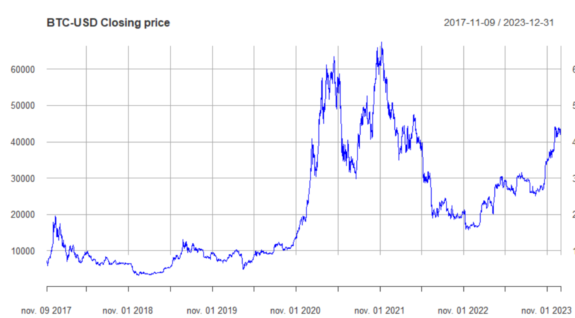


Fig. 1. Evolution of the closing price of BTC ; 2017- 2023

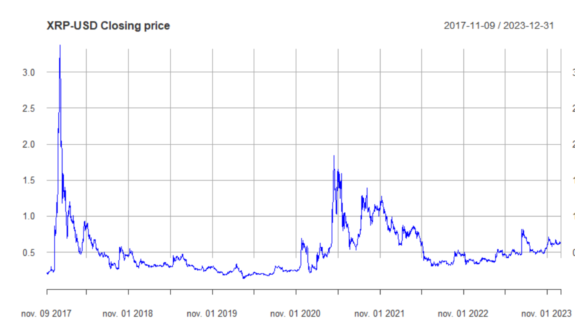


Fig. 2. Evolution of the closing price of XRP; 2017- 2023

## VI. The ARCH test of homoscedasticity

The ARCH (Autoregressive Conditional Heteroskedasticity) test is a statistical method used to identify the presence of conditional heteroscedasticity in time series data. This situation occurs when the variance of the dependent variable

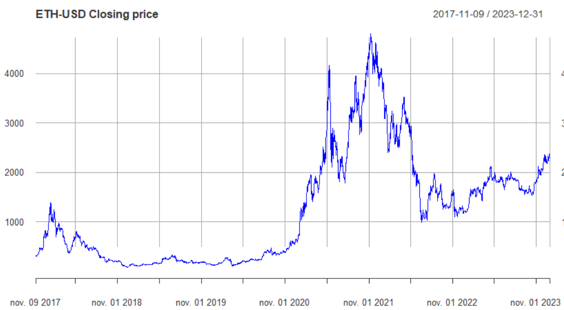


Fig. 3. Evolution of the closing price of ETH; 2017- 2023

changes over time, which is often observed in financial data where the volatility of returns can vary based on new market information. The ARCH test checks the null hypothesis (H0) that the series is homoscedastic, meaning that its variance is constant over time. Rejecting this null hypothesis indicates the presence of conditional heteroscedasticity, indicating that current variations in the series are affected by its own past fluctuations. Conversely, accepting the null hypothesis suggests the absence of ARCH effects, indicating that the variance of errors remains stable over time.

TABLE I
ARCH TEST OF HOMOSCEDASTICITY

| **Serie** | **Chi-Squared** | **df** | **p-value** |
|---|---|---|---|
| BTC. Closing price | 41164 | 20 | $< 0.001$ |
| XRP. Closing price | 19213 | 20 | $< 0.001$ |
| ETH. Closing price | 40711 | 20 | $< 0.001$ |

The table n° I provides the results of ARCH tests on the closing prices for three cryptocurrencies: Bitcoin (BTC), Ripple (XRP), and Ethereum (ETH), revealing significant heteroscedasticity across each one. The high Chi-Squared statistics and p-values less than 0.001 for all three digital currencies reject the null hypothesis of constant variance over time, suggesting that the volatility of closing prices fluctuates and is likely influenced by market information for each data series.

## VII. Conclusion

The results indicate that the volatility of closing prices for Bitcoin (BTC), Ripple (XRP), and Ethereum (ETH) is not constant over time, but instead exhibits significant fluctuations. This rejection of homoscedasticity suggests that simpler models assuming constant volatility may be inadequate for prediction. Instead, these results strongly support the use of more complex volatility prediction models, such as GARCH, EGARCH, or other variants designed to account for the time-varying nature of volatility, as demonstrated by financial time series data. The use of such models has the potential to lead to more accurate forecasts and better risk management strategies in the volatile cryptocurrency market.

## Acknowledgment

I would like to express my deep gratitude to my supervisor, Lamarti Sefian Mohamed, for his valuable guidance and support throughout the duration of this project. His expertise and insightful comments have been crucial in shaping my research, fostering my academic growth, and ensuring the success of this project. I am profoundly grateful to him for his encouragement and the opportunities he has provided me to excel in my field.

## References

[1] Poon, S.-H., & Granger, C. W. J., *Forecasting Volatility in Financial Markets*,Journal of Economic Literature, 41(2), 478-539,2003.

[2] Fleming, J., *IThe Quality of Market Volatility Forecasts Implied by S&P 100 Index* , 1998.

[3] Canina, L., & Figlewski, S., *he Informational Content of Implied Volatility*,Financial Studies, 6(3), 659-681, 1993.

[4] Bollerslev, T., Engle, R. F., & Nelson, D. B., *ARCH Models. In R.F. Engle & D. McFadden (Eds.)*,Vol. 4, 2959-3038. Elsevier Science,1994.

[5] Bjornland, H. C., and Leitemo, K., *dvances in Financial Machine Learning*,A. Wiley,2018.

[6] Lopez de Prado, M., *Identifying the Interdependence between US*, 2009.

[7] Engle, Robert F., *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica, 50 (4): 987-1007,1982.

[8] Bollerslev, T, *Generalised autoregressive conditional heteroskedasticity*,Journal of Econometrics,31, 307-327,1986.

[9] Brooks, C.,*Introductory Econometrics for Finance*, Cambridge University Press,2014.