Financial Market Uncovered – Article 6

Forecasting Volatility: Models, Limits, and Practical Applications

Kilian Voillaume

April 16, 2025

Summary

1 Introduction

An essential component of financial markets is volatility. It gauges the size of changes in asset prices and is essential to almost all facets of quantitative finance, including risk management, regulatory compliance, portfolio creation, and derivatives pricing.

However, volatility is not at all steady. Volatility clustering is the term for the recurrent empirical pattern that may be seen in any financial time series: intervals of calm followed by spikes in turbulence. This demands dynamic models that can capture volatility's time-varying, path-dependent behaviour, challenging the traditional view of volatility as constant or exogenous.

A compelling case study is the financial crisis of 2008. Risk models based on basic historical volatilities significantly underestimated tail risks in the months preceding Lehman Brothers' demise. In the middle of 2008, the S&P 500's one-month implied volatilities were about 20%, but by October, they had risen to over 80%. As volatility erupted and correlations collapsed, firms that relied on static or lagging estimates were taken by surprise and suffered enormous losses. In addition to highlighting the vulnerability of excessively leveraged portfolios, the crisis also demonstrated the shortcomings of models that were unable to predict changes in market regimes.

Forecasting volatility is therefore not just an academic exercise. It is a critical task for:

- Option traders, who need forward-looking volatility estimates to price contracts.
- Risk managers, who must anticipate tail events and compute capital requirements.
- Asset managers, who seek to optimize leverage and rebalance portfolios in response to market stress.

This article reviews key approaches to volatility forecasting, starting with classical models like the *Exponentially Weighted Moving Average (EWMA)* and progressing toward more sophisticated frameworks, such as the *GARCH (Generalized Autoregressive Conditional Heteroskedasticity)* family of models. These models attempt to learn from history — conditioning today's volatility on yesterday's shocks — and account for features like clustering, persistence, asymmetry, and even fat tails.

2 Volatility in financial markets

Volatility is not just a number — it is a dynamic, latent force that influences the pricing of options, the allocation of capital, and the perception of risk. Before building models to forecast it, we need to understand its nature, how it manifests in markets, and what we observe in empirical data. This section introduces key definitions, economic roles, and the statistical patterns that motivate modern volatility modelling.

2.1 What Do We Mean by Volatility?

At its core, volatility refers to the dispersion of returns, typically measured by the standard deviation of logarithmic price changes over a fixed interval:

$$\sigma = \sqrt{\mathbb{E}[(r_t - \mu)^2}$$

This is the definition of volatility as the standard deviation of returns. It tells us how much returns tend to move around their average value. The bigger the average distance from the mean, the more volatile the asset. Think of it as: *'On average, how far does each daily return move away from the typical return?'*

In practice, we distinguish several types of volatility:

Historical (Realized) Volatility

Computed directly from past asset returns. A standard formula for daily log returns over a rolling window of N days is:

$$\sigma_{hist} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_i - \bar{r})^2}$$

It is simple and widely used, but backward-looking and slow to adapt to market changes.

This is the formula for historical volatility: it calculates how spread out the past N returns were from their average. It is a simple way to estimate volatility using a fixed time window (e.g. last 30 days). Think of it as: *'Looking at the last 30 days of returns, how much did they typically deviate from the average return?'*

Implied Volatility (IV)

Implied volatility is the market's consensus estimate of future volatility, embedded in option prices. For a European call option, the Black-Scholes model relates the option price C to the implied volatility σ_{impl}:

$$C = BS(S, K, T, r, \sigma_{impl})$$

Implied volatility varies by strike and maturity, forming a volatility surface. It reflects not just expectations, but also risk aversion, supply/demand, and market frictions.

Forecasted (Conditional) Volatility

Derived from econometric models (e.g., *GARCH* or *EWMA*), conditional volatility is an estimate of the variance of tomorrow's return, given today's information. This forward-looking measure is essential in:

- *VaR* and stress-testing
- Risk-adjusted performance
- Option pricing when market quotes are missing or unreliable

Importantly, realized, implied, and forecasted volatilities differ in how they are derived, what they represent, and how they behave in practice.

2.2 Stylized Facts of Financial Time Series

Volatility is not an independent and identically distributed noise. It follows persistent, nonlinear, and often asymmetric patterns. These patterns — widely observed across markets — are known as stylized facts, and any credible volatility model must reproduce them.

Volatility Clustering

Periods of high volatility tend to follow each other. Mathematically, this is reflected in positive autocorrelation in the squared or absolute returns:

$$Corr(r_t^2, r_{t-k}^2) > 0$$

It implies that large shocks are informative about future risk, and this motivates conditional variance models.

This means that squared returns today are positively correlated with squared returns from the past. In other words, if there was a big move k days ago, there's a good chance today's move is also big — not necessarily in the same direction, but in terms of size.

It captures the idea of volatility clustering: '*When the market is turbulent, it tends to stay turbulent for a while.*'

Mean Reversion

Volatility tends to return to a long-run average over time. After market turmoil, it does not stay high forever — it decays. This property underpins models like *GARCH*, where the conditional variance reverts toward an unconditional (long-run) mean.

$$\mathbb{E}[\sigma_t^2] \rightarrow \sigma_\infty^2$$

This tells us that over time, the forecasted volatility tends to settle around a long-run average — even after short-term shocks.

It is the statistical way of saying: *'Volatility may spike after a crisis, but eventually it calms down and returns to its normal level.'*

Asymmetry (Leverage Effect)

Future volatility is typically more affected by negative returns than by positive returns of the same magnitude. This impact, which is most noticeable in equity markets, implies that volatility depends on both the magnitude and the sign of the return.

Empirical models like *GJR-GARCH* and E*GARCH* account for this by including asymmetric terms in their variance equations.

Fat Tails and Kurtosis

Kurtosis is a statistical metric used to characterize a distribution's form, particularly its peakedness and tails. In contrast to a normal (bell-shaped) distribution, *Kurtosis* indicates the likelihood that a distribution can provide extreme values, such as significant gains or losses. It answers the question: '*How likely is the distribution to yield extreme outcomes?'*

- *Mesokurtic is when kurtosis is roughly 3.* Its moderate tails and lack of an excessive probability of extreme events make it resemble the form of the typical normal distribution.
- *Leptokurtic is when kurtosis exceeds 3.* Its broad tails indicate that it experiences extreme profits or losses more frequently than it would in a typical environment. In the financial markets, this is extremely common.
- *Platykurtic is when kurtosis is less than 3.* Because of its narrow tails, high values are less common than they would be in a normal distribution.

Return distributions are leptokurtic: they exhibit fatter tails and more peakedness than the normal distribution. This means that extreme events happen more often than standard models assume. This undermines any volatility forecast that relies solely on Gaussian assumptions.

$$Kurtosis(r_t) = \frac{\mathbb{E}[(r_t - \mu)^4]}{\sigma^4} \gg 3$$

A normal distribution has a kurtosis of 3. When the kurtosis is much higher than 3, it means the return distribution has fat tails — extreme events (big gains or losses) happen more often than in a normal bell curve.

In practice: *'Markets have more crashes and rallies than a textbook normal model would predict.'*

Persistence and Long Memory

Volatility shocks decay slowly. Autocorrelations in squared returns often remain significant for dozens or hundreds of lags, pointing to long-memory dynamics — sometimes modelled by fractional *GARCH* or *HAR-RV* frameworks.

2.3 Why Stylized Facts Matter

These features are not just statistical curiosities — they have direct implications for trading, hedging, and risk control:

- Underestimating volatility clustering leads to poor risk budgeting.
- Ignoring asymmetry results in ineffective hedging on downside moves.
- Failing to capture fat tails means underpricing options and underpreparing for crises.

A good volatility model is one that is empirically validated, responsive to new information, and robust to structural breaks.

3 Classical models for volatility

Before the development of conditional volatility models like *GARCH*, practitioners used simpler estimators to quantify risk. While these classical models do not fully capture the complex dynamics of financial markets, they remain widely used due to their intuitive appeal and ease of implementation. In this section, we examine two such methods: the rolling window estimator and the *Exponentially Weighted Moving Average (EWMA)* model.

3.1 Rolling window volatility estimator

The rolling window estimator computes volatility as the standard deviation of past returns over a fixed number of days. It assumes that returns are independently and identically distributed within that window.

$$\sigma_{hist} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_i - \bar{r})^2}$$

This formula looks at the last N daily returns (e.g., 30 or 60) and measures how much they tend to vary from their average. It provides a simple snapshot of past volatility.

Example: A 30-day rolling volatility of 20% means that, based on the last month of returns, the asset has shown an annualized standard deviation of 20%.

- **Strengths**

Easy to compute and interpret: It only requires historical return data and basic statistics, making it widely accessible.

No assumptions about return dynamics: It does not impose a specific structure or process, so it is model-free.

- **Limitations**

Equal weighting of all returns: All observations within the window are treated the same, ignoring the fact that more recent returns may carry more information.

Lag in adapting to market changes: Sudden spikes in volatility take time to be fully reflected, especially if the window is large.

Arbitrary window choice: The choice of window length N significantly affects the result and is often chosen arbitrarily (ad hoc).

3.2 *Exponentially Weighted Moving Average (EWMA)*

The *EWMA* model improves upon the rolling estimator by assigning exponentially decreasing weights to past squared returns. This allows the estimate to respond more quickly to recent market movements.

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

Where:

- r_{t-1} is yesterday's return
- σ_{t-1}^2 is yesterday's estimated variance
- $\lambda \in [0,1]$ is the decay factor (e.g., *Risk Metrics* uses $\lambda = 0.94$ for daily data)

The model says that today's volatility is a weighted average of yesterday's volatility and yesterday's squared return — giving more importance to recent data.

If a market shock occurred yesterday, today's volatility estimate will increase more than it would under a rolling window.

- **Strengths**

Responsive to market shocks: By emphasizing recent returns, it quickly adjusts to changes in volatility.

No need to select a fixed window: The decay factor λ controls memory continuously, avoiding abrupt cutoffs.

Simple recursive implementation: It can be updated in real time using only the previous day's estimate and return.

- **Limitations**

A single parameter controls all dynamics: The flexibility of the model is limited to the choice of λ, which may not capture all market regimes.

No long-term volatility level: *EWMA* does not explicitly revert to a long-run variance, which limits its forecasting ability over longer horizons.

No accounting for asymmetry or fat tails: It assumes all returns have equal effect on volatility, regardless of sign or extremity.

4 The *GARCH* model: volatility as a process

Earlier models treat volatility as statistical summary of past returns (*rolling window, EWMA*). However, there are recurring trends in financial markets that imply volatility is a dynamic process that changes over time in response to previous shocks instead of being just a statistic measure.

This realization led to the development of a class of models called *ARCH (Autoregressive Conditional Heteroskedasticity)* and *GARCH (Generalized ARCH)*, in which volatility is directly modelled as a function of historical variances and squared returns.

4.1 The GARCH(1,1) model

The most commonly used specification in practice is the *GARCH*(1,1) model, which assumes that today's variance depends on:

- A long-run average level (the intercept),
- Yesterday's squared return (the shock),
- Yesterday's variance estimate (the persistence).

The model consists of two equations:

$$Return\ equation\colon r_t = \mu + \epsilon_t, \epsilon_t = \sigma_t z_t$$

$$Volatility\ equation\colon \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Where:

- $\alpha_0 > 0$: base volatility (long-run average)
- $\alpha_1 \geq 0$: effect of yesterday's return shock (ARCH term)
- $\beta_1 \geq 0$: persistence of yesterday's volatility (*GARCH* term)
- ϵ_{t-1}^2: is the magnitude of the surprise, squared — it reflects how "big" the shock was, regardless of whether it was positive or negative.
- z_t: $i.i.d.$ standard normal random variable

Today's volatility depends on how big yesterday's surprise was and how volatile the market was yesterday. If the market just had a large move, or was already volatile, we expect today's risk to be high.

The model captures volatility clustering and mean reversion: shocks raise volatility temporarily, but it decays over time.

4.2 Interpretation of parameters

The parameters tell us:

- α_0: The “floor” or long-run volatility level.
- α_1: Sensitivity to market shocks. A high α_1 means volatility reacts strongly to large returns.
- β_1: Volatility persistence. A high β_1 means that once volatility rises, it remains elevated for longer.

For the model to be stationary, the sum $\alpha_1 + \beta_1 < 1$. The closer this sum is to 1, the more persistent the volatility.

4.3 Volatility forecasting with GARCH

One of the major advantages of the GARCH framework is that it treats volatility as a forecastable variable, using an explicit model of its dynamics. Once we estimate the *GARCH* model parameters from historical data, we can generate forward-looking volatility forecasts — essential for risk management, option pricing, and stress testing.

4.3.1 One-step ahead forecast

The *GARCH*(1,1) one-step-ahead forecast simply plugs in the most recent data:

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2$$

Where:

- ϵ_t^2 is the most recent squared return shock
- σ_t^2 is today’s estimated conditional variance

The forecast for tomorrow’s volatility depends on how big the surprise was today (shock) and how volatile the market already was today (persistence).

If a large move just occurred, and today’s volatility is already high, the model will forecast even higher risk for tomorrow.

4.3.2 Multi-step ahead forecasts

To look further ahead — say, 5 or 10 days — we forecast recursively. For a *GARCH*(1,1), the k-step ahead volatility forecast converges toward the long-run average variance:

$$\mathbb{E}[\sigma_{t+h}^2] = \sigma_\infty^2 + (\alpha_1 + \beta_1)^h (\sigma_t^2 - \sigma_\infty^2)$$

Where:

$$\sigma_\infty^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

This formula says that as we look further into the future, the effect of recent shocks fades, and volatility forecasts revert to their unconditional mean.

4.3.3 Interpretation and use cases

Short-term forecasts are highly sensitive to recent shocks. They are useful for:

- Setting intraday or overnight risk limits
- Adjusting portfolio exposure dynamically

Medium-term forecasts gradually smooth out recent noise. They are suitable for:

- *VaR* and *CVaR* estimation over 10–20-day horizons
- Stress testing capital requirements

The speed of mean reversion depends on $\alpha_1 + \beta_1$:

- If close to 1: volatility is highly persistent
- If significantly less than 1: volatility reverts quickly to normal

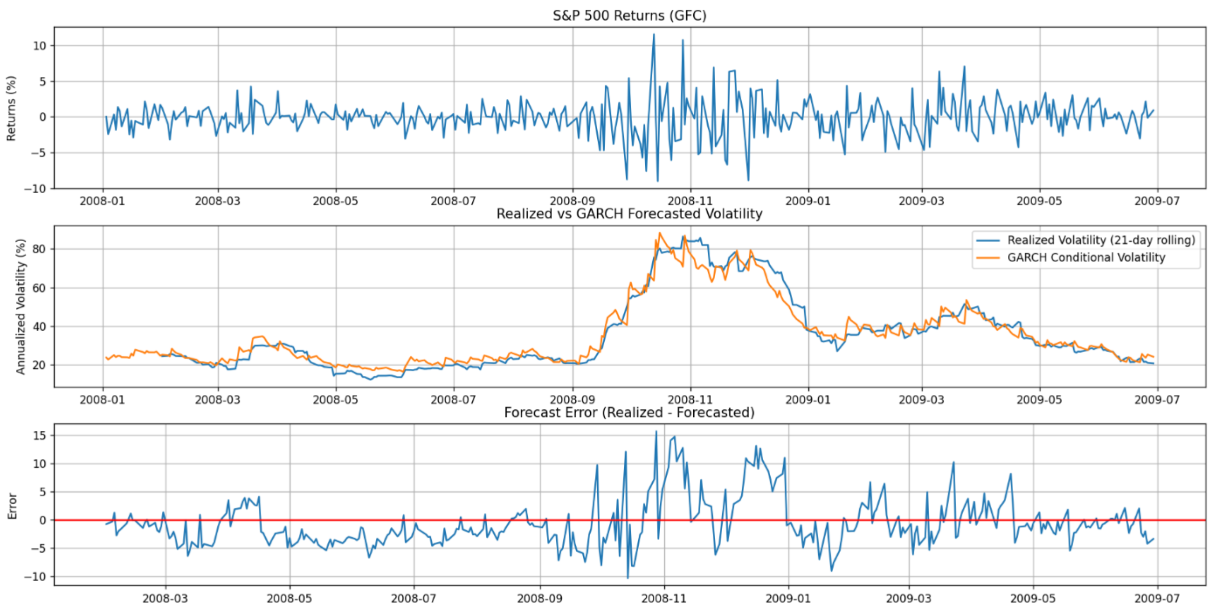

Figure 1: GARCH(1,1) model vs. realized volatility during the 2008 Global Financial Crisis.

This chart illustrates the strength of *GARCH* models during extreme market events. It can be observed that volatility forecasts spike after large negative returns — especially in late 2008 — confirming the model's ability to reflect market stress. However, the delay in forecast adjustment highlights a limitation: *GARCH* reacts but does not anticipate regime changes.

4.4 Strengths and limitations

- **Strengths:**

Captures volatility clustering: Volatility shocks are persistent, as observed in real data — *GARCH* formalizes this process.

Forecasts are model-based: Unlike *EWMA*, *GARCH* treats volatility as a variable governed by its own equation.

Reverts to a long-run mean: Provides more realistic medium-term forecasts.

- **Limitations:**

Assumes symmetric response to shocks: A +5% return affects volatility the same as a -5% return — inconsistent with empirical findings (e.g. leverage effect).

Assumes normality of innovations (model residuals/errors): Fat tails and skewness in real returns are not accounted for in basic *GARCH*.

Linear structure: Volatility reacts in a linear fashion to squared returns, which may miss non-linear effects or jumps.

5 Beyond basic *GARCH*: Asymmetry and Tail Risk

A strong framework for modelling and predicting volatility is offered by the conventional *GARCH*(1,1) model. However, there are several significant drawbacks:

- It treats positive and negative returns as having the same impact on volatility.
- It assumes normally distributed errors, which underestimates the likelihood of extreme returns.

In fact, these constraints are crucial, particularly in equity markets where tail events happen significantly more frequently than Gaussian models indicate, and negative news tends to raise volatility more than positive news. In this part, we examine how the *GARCH* framework might be extended to address these issues.

5.1 Asymmetric GARCH models (capturing leverage effects)

Negative shocks have a greater effect on future volatility than positive shocks of the same size, according to empirical research. This is known as the leverage effect.

Numerous models have been proposed to explain this asymmetry.

5.1.1 GJR-GARCH (Glosten-Jagannathan-Runkle, 1993)

This model adds a term to the *GARCH* equation that captures whether the previous shock was negative:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \cdot 1_{\{\epsilon_{t-1}\}} \cdot \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Where:

- γ_1 measures the additional impact of negative shocks
- $1_{\{\epsilon_{t-1}\}}$ is the indicator function (1 if the shock is negative, 0 otherwise)

If yesterday's return was negative, it increases today's volatility more than a positive return would.

Example: A -5% return leads to a larger increase in volatility than a +5% return — a key feature of real equity markets.

5.1.2 EGARCH (Exponential GARCH, Nelson 1991)

The E*GARCH* model reformulates the volatility equation in logarithmic terms:

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

Features:

- Asymmetry: captured by γ. If $\gamma < 0$, then negative returns raise volatility more than positive ones.

- No need for positivity constraints on parameters (variance is always positive due to the log transformation).
- Captures nonlinear scaling effects better than standard *GARCH*.

Volatility reacts not just to the size of a shock, but also to its sign, and the effect is modelled logarithmically.

5.2 Modelling fat tails

Basic *GARCH* assumes that return shocks ϵ_t are normally distributed — a poor fit for financial data, which shows frequent large deviations from the mean (crashes, surges).

Two common solutions:

- *Student-t GARCH*: assumes $\epsilon_t \sim t_\nu$, where ν is the degrees of freedom. Fatter tails allow for more realistic modeling of extreme events.
- *Generalized Error Distribution (GED):* introduces flexibility in kurtosis by adjusting tail thickness.

Why this matters

- Under normality, *GARCH* underestimates *Value at Risk* (*VaR*).
- Heavy-tailed models provide more accurate forecasts of tail risk and better capture market crashes.

5.3 Example: Market response to a crash

Imagine a sharp -6% drop in equity returns:

- *GARCH*(1,1): Treats it as just a large return — no special impact due to the sign.
- *GJR-GARCH*: Adds an extra volatility shock because the return is negative.
- E*GARCH*: Increases volatility disproportionately due to the negative sign and logarithmic scaling.
- *Student-t GARCH*: Recognizes that such a move is more likely than under the normal distribution — does not view it as a statistical outlier.

6 Real-world applications

Volatility forecasting is not just an academic exercise — it is a cornerstone of quantitative finance. Accurate, forward-looking volatility estimates are essential in:

- Setting trading strategies
- Managing portfolio risk
- Pricing and hedging derivatives
- Meeting regulatory capital requirements

In this section, we explore two core areas of application: risk management and option pricing.

6.1 Volatility forecasting in risk management

6.1.1 Value at Risk – VaR

Volatility forecasts are a direct input into *Value at Risk* models — used by banks, asset managers, and regulators to estimate potential portfolio losses over a given horizon and confidence level.

If we assume portfolio returns are normally distributed, 1-day *VaR* at 99% confidence can be expressed as:

$$VaR_{0.99} = \mu - 2.33 * \hat{\sigma}$$

Where $\hat{\sigma}$ is the forecasted volatility from a *GARCH*, *EWMA*, or other models.

A poor volatility forecast leads to underestimation of potential losses, making portfolios vulnerable to tail events — especially in leveraged environments.

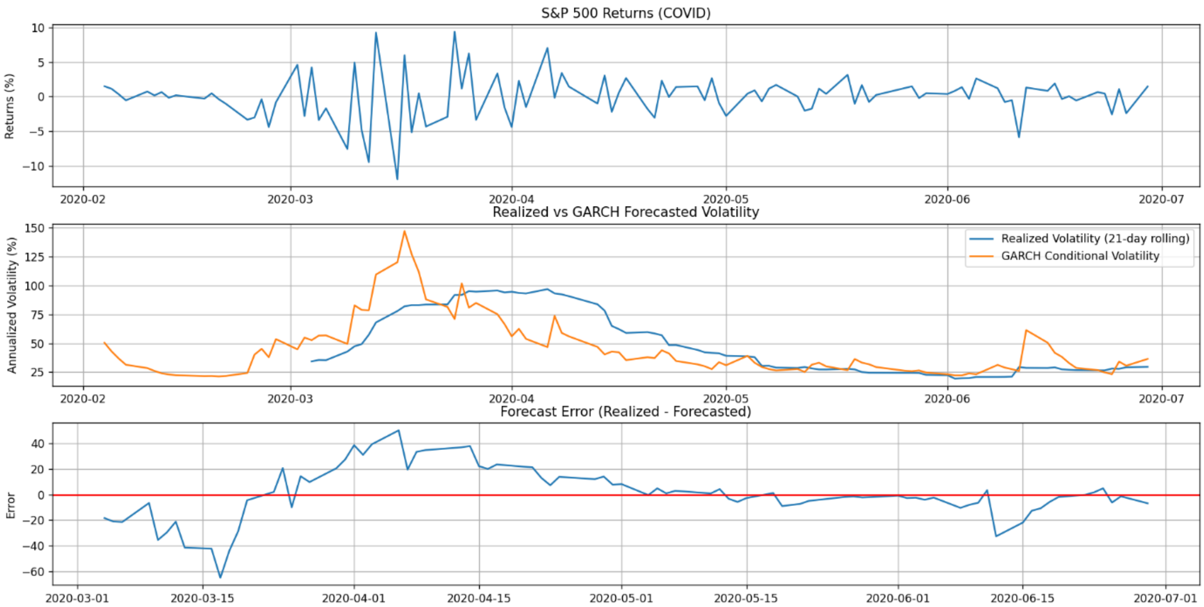

Figure 2: Forecasted vs. realized volatility during COVID-19 (early 2020).

This figure showcases both the strengths and weaknesses of *GARCH* under rapid market collapse. The model recognizes elevated risk, which is essential for *VaR*, but underestimates the full magnitude in real time. This supports the view that *GARCH* should be stress-tested or extended with fat-tailed or regime-switching features when used for risk-sensitive applications during crises.

6.1.2 Conditional VaR and stress testing

Beyond standard *VaR*, Conditional *VaR* (*CVaR*) — the expected loss beyond the *VaR* threshold — is also based on volatility forecasts. Models like Student-t *GARCH* provide more accurate tail risk estimates than Gaussian models.

Regulators (e.g. under Basel III/IV) often require stressed *VaR*, which uses volatility models calibrated during periods of market turmoil.

6.1.3 Dynamic risk budgeting

Funds using volatility-based allocation (e.g. volatility targeting, risk parity) adjust their exposure based on expected risk. When the forecasted volatility increases, portfolio weights are reduced to maintain a constant risk profile.

6.2 Volatility in derivatives pricing

6.2.1 Option pricing models

Theoretical models like Black-Scholes require volatility as an input. While implied volatility is often used in practice, in the absence of liquid options, practitioners rely on forecasted volatility from *GARCH* or *EWMA*.

In Black-Scholes, the price of a European call is:

$$C = BS(S, K, T, r, \hat{\sigma})$$

Where $\hat{\sigma}$ is the volatility forecast over the option's life. For short-dated options, a 1-day or 5-day *GARCH* forecast is scaled to the option horizon.

For illiquid markets (emerging assets, crypto, long-dated OTCs), model-based volatility becomes a critical input.

6.2.2 Volatility forecast vs Implied volatility

Comparing forecasted and implied volatility offers a powerful trading signal:

- If implied volatility is significantly higher than forecasted volatility → Sell volatility (e.g. sell options)
- If implied volatility is lower → Buy volatility

This forms the basis of many volatility arbitrage strategies.

6.3 Other practical applications

Hedging ratios: Volatility forecasts help determine hedge ratios in dynamic delta hedging or minimum-variance strategies.

Limit setting: Risk teams adjust trading limits based on forecasted volatility, tightening constraints during stress.

Performance attribution: Volatility-adjusted performance metrics (e.g., Sharpe ratio) rely on conditional variance estimates

6.4 Comparative summary

Use case	*Model input*	*Why volatility matters*
VaR/CVaR	GARCH, EWMA	Determines scale of potential losses
Option pricing	Forecasted σ	Essential for fair value estimation when IV is unreliable
Risk targeting strategies	Forecasted σ	Drives exposure deicisions in volatility-managed portfolios
Volatility arbitrage	Forecasted vs IV	Signal for under/overpricing of option
Regulatory capital	Stressed GARCH	Required for capital reserves and stress tests

7 Limitations and extensions of the *GARCH* framework

Despite its widespread use and empirical success, the *GARCH* family is not a silver bullet. While it models volatility clustering and persistence effectively, it falls short in several key areas — especially when applied in complex or turbulent market environments.

In this section, we discuss the limitations of classical *GARCH* models and explore modern extensions designed to improve their performance and realism.

7.1 Structural limitations of GARCH

Linearity in Squared Returns:

GARCH assumes that volatility responds linearly to squared shocks. But in practice, market reactions to large events (crashes, geopolitical surprises) may be nonlinear and regime-dependent.

For instance, a -7% return may trigger disproportionately more volatility than predicted by a linear model.

Lack of Structural Shifts:

GARCH models are stationary by design. They assume that parameters (e.g., shock sensitivity and persistence) are constant over time. This is problematic in the presence of structural breaks — such as central bank regime changes, financial crises, or COVID-like shocks.

In a world of structural change, a model assuming fixed parameters is, by definition, miscalibrated.

Underperformance in Tail Risk Modelling:

Although *GARCH* captures conditional heteroskedasticity, standard versions with normal innovations underestimate extreme risk. Even t-distributed extensions may struggle with true jump behaviour or sudden regime transitions.

7.2 Practical challenges in implementation

Parameter estimation instability:

GARCH models are sensitive to outliers and initial values. In small samples, estimation via maximum likelihood can be noisy or fail to converge.

Scaling issues for portfolio-level modelling:

In multivariate settings, the number of parameters grows quickly (quadratically), making full covariance matrix estimation unstable.

Misspecification risk:

Selecting the wrong model variant (e.g., *GARCH* vs E*GARCH* vs *GJR-GARCH*) can lead to poor forecasts. There is no single universally best choice.

7.3 Extensions and alternatives

As markets have become more complex, so have the models we use to understand volatility. Below are four popular extensions of *GARCH* that help capture behaviours *GARCH* can not explain — such as sudden shifts in market conditions, different investor behaviours, or unexpected crashes.

7.3.1 SV models – Stochastic volatility models

Instead of modelling volatility as a deterministic function of past returns, SV models treat volatility itself as an unobservable stochastic process:

$$r_t = \mu + \sigma_t z_t, \log(\sigma_t^2) = \alpha + \beta * \log(\sigma_{t-1}^2) + \eta_t$$

This model assumes that volatility itself is unpredictable, just like returns. It moves randomly over time, without being directly tied to yesterday's return.

In *GARCH*, volatility is calculated directly from previous return shocks. But in real markets, volatility sometimes changes even when returns are calm — for example, when uncertainty grows because of a central bank announcement or a war. Stochastic volatility models allow volatility to change on its own, based on its own random path. They are especially useful when pricing options or dealing with products that are sensitive to volatility dynamics (like volatility swaps), but they are harder to estimate and interpret than *GARCH*.

7.3.2 MS-GARCH – Markov-Switching GARCH

Markets alternate between regimes (e.g., calm and crisis). MS-*GARCH* models allow the parameters of the *GARCH* equation to change depending on a latent state that follows a Markov chain:

$$\sigma_t^2 = \alpha_{0,S_t} + \alpha_{1,S_t}\epsilon_{t-1}^2 + \beta_{1,S_t}\sigma_{t-1}^2$$

This model assumes the market is in one of several hidden "states" (like calm or crisis), and each state has its own volatility behaviour. The market switches between states randomly over time.

Think of the market as having moods — sometimes it is calm, other times it is panicked. MS-*GARCH* models assume that each mood has its own volatility rules. For example, in calm periods, volatility reacts mildly to news; in crises, it reacts strongly and stays high. These models are good at capturing sudden market shifts, like the 2008 crash or COVID-19 panic. However, they are more complex and harder to work with in real time because we do not know for sure which state we are in until after the fact.

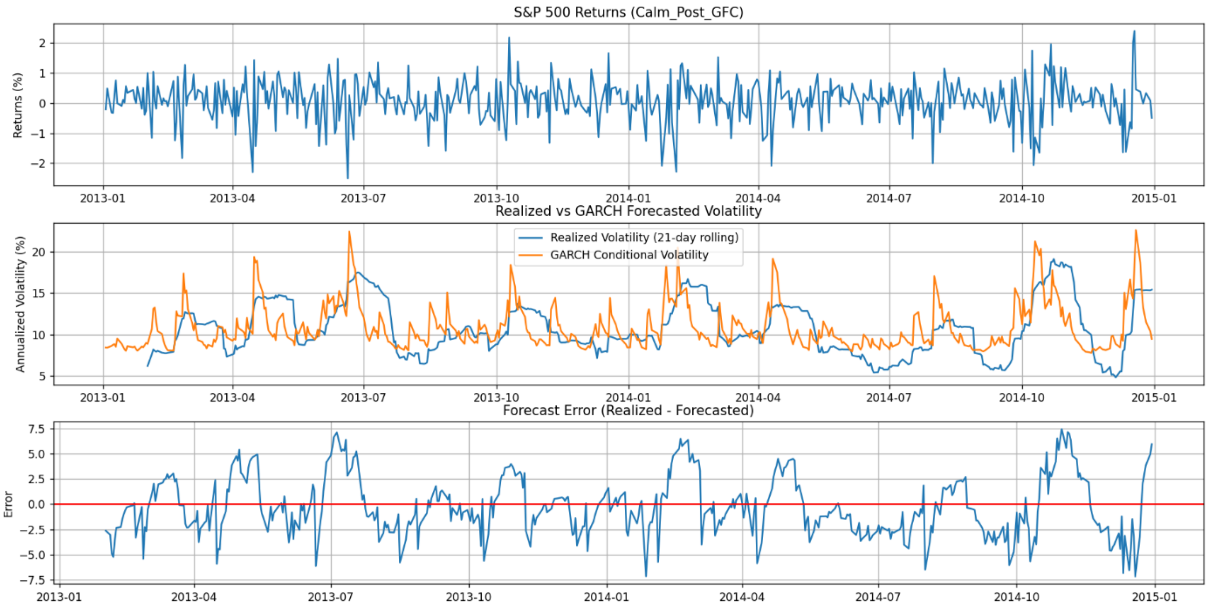

Figure 3: GARCH(1,1) model during the 2013–2014 post-crisis calm period.

In calm regimes, *GARCH* tends to overreact to isolated return shocks and overpredict volatility. This overestimation creates noise in *VaR* or margin settings and confirms that *GARCH*'s linear structure may be too sensitive in low-volatility markets. This motivates the use of smoother or regime-sensitive models (e.g. *HAR-RV*, MS-*GARCH*) for such environments.

7.3.3 HAR-RV – Heterogeneous Autoregressive Realized Volatility

Proposed by Corsi (2009), *HAR-RV* models realized volatility using multiple time horizons (daily, weekly, monthly):

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1}^{week} + \beta_3 RV_{t-1}^{month} + \epsilon_t$$

This model predicts volatility using volatility from different timeframes — yesterday, last week, and last month — because different types of investors watch different time horizons.

Not all market participants behave the same way. A day trader looks at yesterday's volatility. A pension fund looks at last month's. This model combines those views to forecast future volatility more accurately. It performs well in capturing realized volatility — the actual movement we see — especially when using high-frequency data (e.g., 5-minute returns). It is also simple to use and often outperforms *GARCH* in forecasting contests. But it does not explain why volatility moves — it just learns patterns from the data.

7.3.4 Machine Learning Models

These models let the computer learn patterns in the data by itself, without assuming a specific equation. They use past returns, volatility, and other variables to predict future volatility.

Machine learning models are flexible and powerful. They can detect complex, hidden relationships in the data that standard models miss. For example, they might learn that certain types of news or macro indicators often lead to higher volatility — even if returns do not move much right away. These models are increasingly used by hedge funds and algorithmic traders. But they are often black boxes: we don't always know why they work, and they can fit the training data too closely and perform poorly on new data if not handled carefully. They are great when predictive accuracy matters more than interpretability.

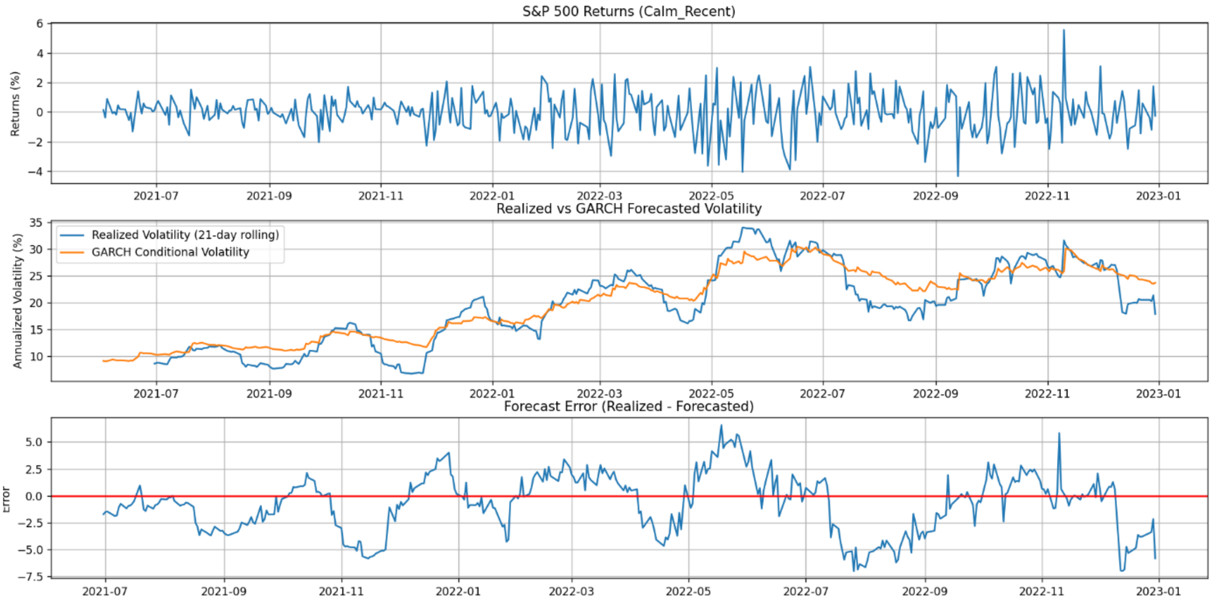

Figure 4: GARCH performance in the recent macro-volatility phase (2021–2022)

In recent markets, volatility has been episodic but hard to model. *GARCH* captures the broad trend but fails to capture transient volatility spikes (e.g., around Fed announcements). This highlights the limits of memory-based models and supports the adoption of more adaptive frameworks, such as *HAR-RV* or *machine learning (ML)*, for volatility forecasting in unpredictable macro environments.

7.4 Summary

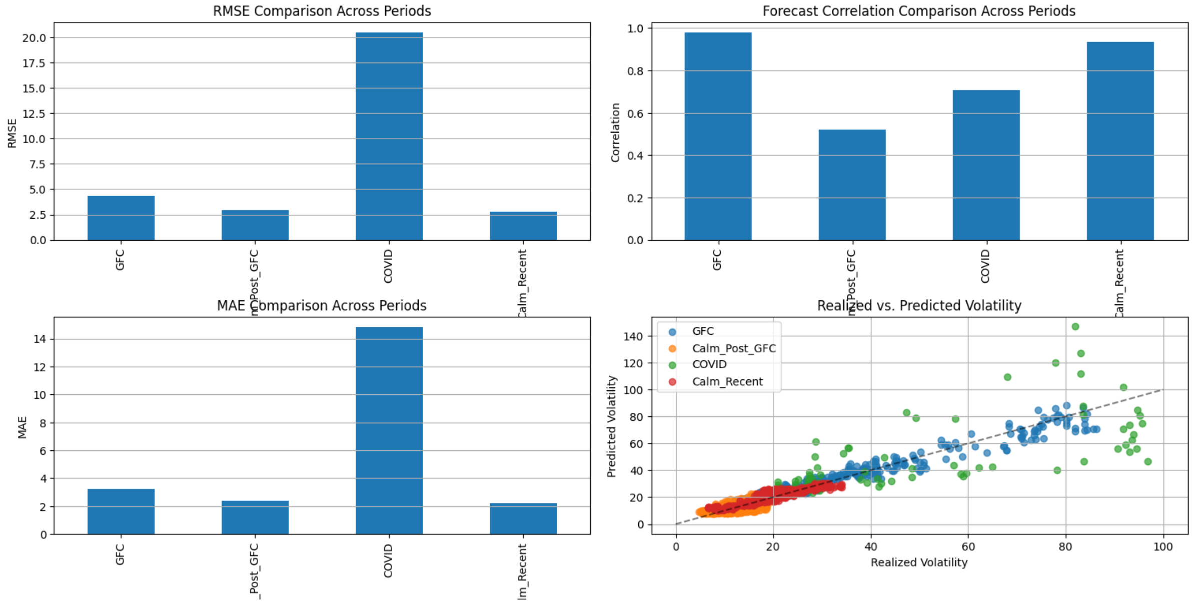

Figure 5: Comparative Forecast performance across market regimes

This figure is the quantitative summary of the article's findings: *GARCH* is more accurate when volatility is elevated and persistent. In stable markets, its forecast error increases and the correlation with realized volatility decreases, supporting the use of regime-adaptive or mixed models in practice.

8 Conclusion

The foundation of financial decision-making is volatility forecasting. Accurate projections of future volatility are crucial for managing uncertainty, whether they are used to price derivatives, distribute money, control risk exposure, or gauge market sentiment.

From basic tools like rolling averages and *EWMA* to more complex frameworks like *GARCH* and its expansions, we have looked at a variety of models in this article. Every model represents a distinct perspective on market behaviour, and each involves trade-offs between empirical accuracy, adaptability, and simplicity.

Classical *GARCH* models have become industry standards for a reason: they capture volatility clustering, persistence, and mean reversion in a relatively tractable framework. However, they are not without limitations. They often assume symmetric responses to market shocks and struggle to adapt to structural breaks, tail risk, or nonlinear behaviours.

To address these shortcomings, practitioners have turned to richer models:

- Asymmetric *GARCH* variants account for the empirical fact that volatility tends to increase more after negative shocks.
- Stochastic volatility models introduce randomness in volatility itself, providing a closer fit to option markets.
- Markov-switching and regime-based models acknowledge that financial markets operate under different states, with abrupt transitions.
- Realized volatility models like *HAR-RV* offer practical advantages when high-frequency data is available.
- Finally, machine learning approaches are gaining traction, offering flexible, data-driven forecasts — though often at the cost of interpretability.

There is no universally superior model. The choice depends on the application, the data at hand, and the level of interpretability or theoretical structure required.

Ultimately, volatility forecasting is not just about selecting a model — it is about understanding risk in a probabilistic, dynamic, and forward-looking way. In a financial world where uncertainty is constant, robust volatility modelling remains one of the most important quantitative tools available to practitioners and researchers alike.

9 Python code

```python
import pandas as pd
import numpy as np
from arch import arch_model

def fit_garch_model(returns, p=1, q=1, forecast_horizon=1):
    returns = returns.dropna()
    model = arch_model(returns, vol='Garch', p=p, q=q, rescale=False)
    model_fit = model.fit(disp='off')

    # Conditional volatility (annualized)
    conditional_vol = pd.Series(
        model_fit.conditional_volatility * np.sqrt(252),
        index=returns.index
    )

    return model_fit, conditional_vol

def calc_realized_volatility(returns, window=21):
    """21-day rolling standard deviation of returns, annualized"""
    return returns.rolling(window=window).std() * np.sqrt(252)
```

Figure 6: Fitting the GARCH(1,1) model and getting the realized volatility estimation

```python
from sklearn.metrics import mean_squared_error, mean_absolute_error
from scipy.stats import pearsonr

def evaluate_forecasts(realized, forecasted):
    joined = pd.concat([realized, forecasted], axis=1).dropna()
    realized_aligned = joined.iloc[:, 0]
    forecasted_aligned = joined.iloc[:, 1]

    mse = mean_squared_error(realized_aligned, forecasted_aligned)
    rmse = np.sqrt(mse)
    mae = mean_absolute_error(realized_aligned, forecasted_aligned)
    corr = pearsonr(realized_aligned, forecasted_aligned)[0]

    return {
        'MSE': mse,
        'RMSE': rmse,
        'MAE': mae,
        'Correlation': corr
    }
```

Figure 7: Comparing performance across regimes

```python
from arch import arch_model

# One period's returns:
gjr_model = arch_model(returns, vol='GARCH', p=1, q=1, o=1)
gjr_fit = gjr_model.fit(disp='off')

# Get asymmetry coefficient
asymmetry_coef = gjr_fit.params.get('gamma[1]', None)
p_value = gjr_fit.pvalues.get('gamma[1]', None)

print(f"Asymmetry coefficient: {asymmetry_coef:.4f}")
print(f"Significant at 5%: {'Yes' if p_value < 0.05 else 'No'}")
```

Figure 8: Testing asymmetry (GJR-GARCH)

10 References

[1] Hull, J. C. (2018). Options, Futures, and Other Derivatives (10th ed.). Pearson.

[2] Alexander, C. (2008). Market Risk Analysis II: Practical Financial Econometrics. John Wiley & Sons.

[3] Alexander, C. (2008). Market Risk Analysis I: Quantitative Methods in Finance. John Wiley & Sons.

[4] Wilmott, P. (2006). Paul Wilmott on Quantitative Finance (2nd ed.). John Wiley & Sons.

[5] Wilmott, P. (2007). Paul Wilmott Introduces Quantitative Finance (2nd ed.). John Wiley & Sons.

[6] Taleb, N. N. (1997). Dynamic Hedging: Managing Vanilla and Exotic Options. John Wiley & Sons.

[7] Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987–1007.

[8] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31(3), 307–327.

[9] Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). 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.

[10] Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, 59(2), 347–370.

[11] Taylor, S. J. (1986). Modelling Financial Time Series. John Wiley & Sons.