ChatGPT

# ARCH vs GARCH for Bursty Volatility and Volatility Clustering

## Executive summary

Volatility clustering—periods of high volatility followed by high volatility, and low followed by low—is a central "stylised fact" of financial returns and other economic time series. [1] Bursty volatility is a closely related description that emphasises *intermittent spikes* ("irregular bursts") in volatility measures. [2]

Both **ARCH** and **GARCH** are *designed* to model **time-varying conditional variance**, which is the formal route by which they represent volatility clustering and burstiness. [3] Historically and technically:

- **ARCH (Engle, 1982)** introduced the core idea: model the **one-step-ahead forecast variance** as a function of *recent past squared forecast errors*, rather than treating variance as constant. [4] This directly targets the empirical regularity that squared residuals (and other volatility proxies) are serially dependent even when raw returns are not. [5]
- **GARCH (Bollerslev, 1986)** was developed as a **generalisation** that includes **lagged conditional variances** in the variance equation. This yields a much more flexible (and typically more persistent) volatility dynamic with far fewer parameters than a high-order ARCH. [6] Bollerslev explicitly motivated GARCH as accommodating the "long memory" *typically found in empirical work* and providing a "more parsimonious description." [7]

**Which model is *designed for volatility clustering and bursty volatility?***
***The clean answer is: ARCH is the foundational model introduced to model changing conditional variance (and thereby clustering), while GARCH was specifically developed to capture persistent clustering / burstiness more parsimoniously and realistically*** by allowing volatility to depend on its own past (and by producing slowly decaying effects via $\alpha + \beta$ persistence). [8]

Empirically, even in Bollerslev's original paper, a simple **GARCH(1,1)** is argued to provide a **(slightly/marginally) better fit** and a **more plausible learning mechanism** than an **ARCH** with an ad hoc long lag structure. [9] For "bursty" behaviour that goes beyond geometric persistence (e.g., extremely long memory, asymmetry/leverage, or jump-like spikes), standard GARCH often needs **extensions** (EGARCH, GJR-GARCH, FIGARCH, heavy-tailed innovations, regime-switching, etc.). [10]

## Foundations and definitions

ARCH and GARCH models are built around a standard decomposition for an observed series (often returns) into a conditional mean and an innovation whose scale changes over time. A common representation is:

$$r_t = \mu_t + a_t, \qquad a_t = \sigma_t \varepsilon_t, \qquad \mathbb{E}(\varepsilon_t) = 0,\ \mathbb{V}(\varepsilon_t) = 1,$$

where $\sigma_t^2$ is the **conditional variance** (the modelled volatility). [11]

### ARCH definition

In **ARCH**, the conditional variance is a function of *lagged squared innovations*. A canonical ARCH($q$) specification is:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i a_{t-i}^2, \quad \omega > 0,\ \alpha_i \geq 0.$$

This makes the one-step forecast variance depend on "recent past" information (squared forecast errors). [12] In the accessible abstracted framing of Robert F. Engle [13] 's original contribution, ARCH processes have variances that are **nonconstant conditional on the past** but often treated as **constant in an unconditional sense**, which is exactly the conceptual shift away from constant-variance time series models. [14]

### GARCH definition

**GARCH** augments ARCH by allowing lagged conditional variances to enter the variance recursion:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \quad \omega > 0,\ \alpha_i, \beta_j \geq 0.$$

This structure is explicitly presented as the "natural generalisation" that includes past conditional variances. [15]

In many applications, **GARCH(1,1)** dominates for parsimony:

$$\sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

with a standard stationarity condition (in many textbook treatments) $\alpha_1 + \beta_1 < 1.$ [16]

## Formal mechanisms for time-varying volatility and clustering

### Why ARCH generates clustering

ARCH creates volatility clustering through a **feedback loop from shocks to future variance**. If $a_{t-1}^2$ is large, the model sets $\sigma_t^2$ larger; that larger $\sigma_t$ increases the scale of $a_t$, making another large $a_t^2$ more likely, and so on for as many lags as are included. This is the model's direct translation of "squared returns are autocorrelated" into a parametric dynamic law. [17]

But in plain ARCH($q$), persistence is limited by the explicit lag order $q$: if the data exhibit long-lived volatility dependence, ARCH often needs a relatively large $q$, which creates practical constraints and estimation burdens. [18]

### Why GARCH generates (and intensifies) clustering

GARCH adds an additional channel: **volatility depends on its own past**. This means shocks affect volatility today via $a_{t-1}^2$, and volatility itself carries forward via $\sigma_{t-1}^2$. The result is that volatility shocks can decay **geometrically** at a rate essentially governed by $\alpha_1 + \beta_1$ in the simplest case, which is exactly why GARCH became the workhorse for persistent clustering. [19]

Two primary-source statements capture the design intent:

- Tim Bollerslev [20] motivates GARCH as addressing "the long memory typically found in empirical work" and providing "a more parsimonious description." [7]
- Robert F. Engle [13] (Nobel lecture) describes GARCH as generalising ARCH "to an autoregressive moving average model," with weights on past squared residuals "assumed to decline geometrically." [21]

Textbook exposition makes the clustering mechanism explicit: in a GARCH(1,1), "a large $a_{t-1}^2$ ... tends to be followed by another large $a_t^2$, generating ... volatility clustering." [22]

## Volatility clustering and bursty volatility

### What "volatility clustering" means (formal and empirical)

A widely used empirical definition is: volatility clustering is the phenomenon that "large price variations are more likely to be followed by large price variations." [23] It is commonly operationalised by positive serial dependence in volatility proxies such as $|r_t|$, $r_t^2$, or realised/smoothed volatility measures; in Cont's formulation, clustering is measured via autocorrelation of squared returns. [23]

ARCH/GARCH are explicitly described as designed to address the fact that practitioners "describe this as 'volatility clustering'," and that these models "are designed to deal with just this set of issues." [24]

### What "bursty volatility" adds beyond clustering

"Bursty" volatility typically emphasises **intermittency**—quiet periods punctuated by sharp spikes—rather than only persistence. Cont's stylised-fact discussion connects this directly to "irregular bursts" in volatility estimators as a hallmark of variable volatility. [25]

ARCH/GARCH can represent burstiness because a single large shock raises $\sigma_t^2$ immediately, creating a "burst", and then subsequent decay produces a cluster whose *duration* depends on persistence ($q$ in ARCH, $\alpha + \beta$ in GARCH). [26]

However, for *isolated* spikes (large shocks not followed by sustained turbulence), basic ARCH can behave poorly: a standard critique is that "ARCH models are likely to overpredict the volatility because they respond slowly to large isolated shocks." [27] This is one practical sense in which "bursty" data can stress pure ARCH specifications.

## Which model was developed to capture clustering, and evidence on bursty performance

### Development motivation: from ARCH to GARCH

ARCH's introduction is fundamentally about rejecting constant forecast variance. In the accessible abstract of the reprinted Engle contribution, "Traditional econometric models assume a constant one-period forecast variance" and ARCH processes are introduced so that the "recent past gives information about the one-period forecast variance." [28]

GARCH's development is explicitly framed as enabling **longer memory and a more flexible lag structure** than is typically practical in finite-order ARCH—directly aimed at persistent clustering patterns observed empirically. [6] In other words:

- **ARCH**: designed to make volatility time-varying and conditionally predictable from the recent past. [4]
- **GARCH**: designed to make that time variation **persist realistically** and **parsimoniously** (capturing clustering/bursts without large $q$). [6]

### Evidence and comparisons focused on bursty/clustering behaviour

Within the original GARCH paper, Bollerslev provides directly comparative evidence: a simple GARCH model "provides a marginally better fit" and "a more plausible learning mechanism" than an ARCH model with an ad hoc, relatively long lag structure. [9] This is especially relevant to "bursty clustering", because the comparison is precisely about how volatility responds across time to shocks (i.e., the implied lag/weight structure). [29]

A second type of evidence is practical forecasting success: the JEP exposition notes that the GARCH generalisation has "proven surprisingly successful in predicting conditional variances," and highlights its slowly declining weights that "never go completely to zero." [30] These statements are aligned with the empirical reality that volatility clustering lasts longer than a small finite number of lags in many datasets. [31]

Finally,Cont's stylised facts emphasise that even once clustering is modelled (e.g., "via GARCH-type models"), data may retain "conditional heavy tails," signalling a limitation of variance-only dynamics for extreme burstiness unless the innovation distribution and/or model class is enriched. [32]

## Limitations and typical extensions for burstiness

### Core limitations of plain ARCH and baseline GARCH

Three limitations are especially important when the practical goal is capturing *bursty* volatility realistically:

First, **symmetry**: because variance depends on squared shocks, a basic ARCH treats positive and negative shocks as having the same volatility effect, whereas financial data often show asymmetry ("leverage effects"). [33]

Second, **isolated bursts and overreaction/drag**: textbook discussion highlights that ARCH can "respond slowly to large isolated shocks," leading to volatility overprediction after a one-off burst. [27] (This is a concrete "bursty volatility" failure mode: difficulty separating a true regime of high volatility from a single outlier.)

Third, **residual tail risk after volatility scaling**: even after adjusting for volatility clustering using GARCH-type models, heavy tails may remain, implying that burstiness is not only about conditional variance—it is also about the conditional distribution of shocks. [34]

### Extensions commonly used for clustering and burstiness

The following extensions are standard responses to these limitations, each targeting a different aspect of burstiness:

**Asymmetric GARCH family (leverage and sign effects).** Engle's Nobel lecture highlights that EGARCH-type models recognise that volatility can respond "asymmetrically" and that in financial contexts "negative returns seemed to be more important predictors of volatility than positive returns." [21] This motivates EGARCH and closely related forms (e.g., GJR-GARCH/TGARCH) used to reproduce asymmetry in bursts.

**Long-memory volatility models (very persistent clustering).** If burstiness includes very slow decay of volatility dependence, FIGARCH is a canonical extension. Its abstract states that FIGARCH implies a "slow hyperbolic rate of decay" in the influence of lagged squared innovations—precisely a mechanism for very long-lived clustering beyond geometric decay. [35]

**Heavy-tailed and skewed innovations (extreme bursts).** A substantial share of "bursty" behaviour is distributional: large jumps in returns occur more often than Gaussian shocks. The need for richer innovation distributions is implicit in both the "conditional heavy tails" point and in textbook treatments that discuss heavier tails and kurtosis properties under GARCH specifications. [36] Common practice is to fit GARCH with Student-$t$, GED, or skewed distributions, sometimes combined with asymmetry terms.

**Regime-switching and stochastic volatility (structural bursts).** For bursts that look like discrete regime changes (e.g., crisis vs calm), practitioners frequently use regime-switching volatility models or stochastic volatility (SV) models. While formally distinct from ARCH-family "observation-driven" models, they are often treated as competing approaches for bursty dynamics. [37]

## Comparison table, timeline, and further reading

### Comparison table: suitability for volatility clustering and burstiness

| Attribute | ARCH (Engle 1982) | GARCH (Bollerslev 1986) |
|---|---|---|
| Core goal | Make forecast variance time-varying and conditionally predictable from past squared shocks [4] | Extend ARCH to include past conditional variances; capture persistence more flexibly and parsimoniously [38] |
| Conditional variance equation (typical) | $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i a_{t-i}^2$ [27] | $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$ [39] |
| Mechanism for clustering | Finite-lag feedback from shocks to variance; clusters last roughly over the included lag window [40] | Shock + variance persistence; geometric decay (often via $\alpha + \beta$) yields longer-lived clusters [41] |
| "Bursty volatility" (intermittent spikes) | Creates bursts when $a_{t-1}^2$ is large, but may overpredict after isolated spikes due to slow adjustment [42] | Creates bursts and sustains them (if persistent); often more realistic for clustered episodes; still may miss jump-driven spikes without extensions [43] |

| Attribute | ARCH (Engle 1982) | GARCH (Bollerslev 1986) |
|---|---|---|
| Parsimony | Often needs relatively high $q$ for persistent clustering [18] | Achieves long persistence with low order (e.g., GARCH(1,1)); "more parsimonious description" [44] |
| Evidence in original GARCH paper | — | GARCH argued to provide "marginally" / "slightly" better fit and more plausible lag structure vs high-order ARCH in an inflation-uncertainty example [9] |
| Key limitations for burstiness | Symmetry; parameter constraints; overprediction after isolated shocks [27] | Symmetry in baseline form; geometric decay may be too short vs true long memory; heavy tails may remain after "GARCH-type" scaling [45] |
| Typical extensions | High-order ARCH, ARCH-M, etc. (often supplanted by GARCH variants) [46] | EGARCH / GJR for asymmetry; t-innovations; IGARCH/FIGARCH for persistence/long memory; multivariate DCC-GARCH, etc. [47] |

### Simple mermaid timeline of model development

```
timeline
  title Conditional heteroskedasticity models for clustered/bursty volatility
  1982 : Engle introduces ARCH (time-varying conditional variance)
  1986 : Bollerslev introduces GARCH (adds lagged conditional variance;
parsimonious persistence)
  2001 : Engle popular exposition of ARCH/GARCH and volatility clustering
(JEP)
  2003 : Engle’s Nobel lecture summarises ARCH framework and extensions
  1996 : Baillie–Bollerslev–Mikkelsen introduce FIGARCH (hyperbolic/long-
memory decay)
```

### Recommended further reading

Primary and highly authoritative starting points:

- Robert F. Engle [13] 's ARCH introduction (reprinted abstract accessible) [28] and Nobel lecture overview of ARCH and its extensions [48]
- Tim Bollerslev [20] 's original GARCH paper (including discussion of long memory, parsimony, and comparative fit vs high-order ARCH) [9]
- Analysis of Financial Time Series [49] (conditional heteroskedasticity chapter: equations, stationarity, clustering explanation, practical weaknesses) [50]
- Rama Cont [51] on stylised facts (volatility clustering definition; "irregular bursts"; conditional heavy tails after GARCH-type scaling) [52]
- James D. Hamilton [53] 's discussion of ARCH relevance in macroeconomics applications (context and motivation) [54]
- FIGARCH (long-memory volatility persistence) abstract statement of hyperbolic decay mechanism [35]

### Full citations and links (DOI / stable-access pointers)


```
Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with
Estimates of the Variance of United Kingdom Inflation.
  DOI (JSTOR): 10.2307/1912773
  Reprinted chapter abstract on Oxford Academic: https://doi.org/10.1093/oso/
9780198774310.003.0001

Bollerslev, T. (1986). Generalized Autoregressive Conditional
Heteroskedasticity. Journal of Econometrics, 31, 307–327.
  Publisher PDF (author page): https://public.econ.duke.edu/~boller/
Published_Papers/joe_86.pdf

Engle, R. F. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied
Econometrics. Journal of Economic Perspectives, 15(4), 157–168.
  PDF copy: https://www.fsb.miamioh.edu/lij14/672_engle.pdf

Engle, R. F. (Nobel lecture PDF). Risk and Volatility: Econometric Models and
Financial Practice (overview of ARCH/GARCH and extensions).
  PDF: https://www.nobelprize.org/uploads/2018/06/engle-lecture.pdf

Cont, R. (2001). Empirical properties of asset returns: stylized facts and
statistical issues. Quantitative Finance, 1(2), 223–236.
  DOI: 10.1080/713665670
  (Accessible full-text copy often mirrored via academic hosts; one gateway
page used here was ResearchGate.)

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally
integrated generalized autoregressive conditional heteroskedasticity. Journal
of Econometrics, 74(1), 3–30.
  DOI: 10.1016/S0304-4076(95)01749-6
  ScienceDirect abstract: https://www.sciencedirect.com/science/article/abs/
pii/S0304407695017496

Hamilton, J. D. (2007/2008). Macroeconomics and ARCH (working paper).
  PDF: https://econweb.ucsd.edu/~jhamilto/JHamilton_Engle.pdf
```

---

1 2 23 25 31 32 34 36 45 52 (PDF) Empirical properties of asset returns: Stylized facts and statistical issues
https://www.researchgate.net/publication/313727065_Empirical_properties_of_asset_returns_Stylized_facts_and_statistical_issues

3 4 8 12 14 28 Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation | Arch: Selected Readings | Oxford Academic
https://academic.oup.com/book/52277/chapter/421087438

5 10 13 17 19 21 37 40 41 46 47 48 53 nobelprize.org
https://www.nobelprize.org/uploads/2018/06/engle-lecture.pdf

6 7 9 15 18 29 38 43 44 49 51 public.econ.duke.edu
https://public.econ.duke.edu/~boller/Published_Papers/joe_86.pdf

11 16 22 26 27 33 39 42 50 cpb-us-w2.wpmucdn.com
https://cpb-us-w2.wpmucdn.com/blog.nus.edu.sg/dist/0/6796/files/2017/03/analysis-of-financial-time-series-copy-2ffgm3v.pdf

20 35 Fractionally integrated generalized autoregressive conditional heteroskedasticity - ScienceDirect
https://www.sciencedirect.com/science/article/abs/pii/S0304407695017496

24 30 fsb.miamioh.edu
https://www.fsb.miamioh.edu/lij14/672_engle.pdf

54 econweb.ucsd.edu
https://econweb.ucsd.edu/~jhamilto/JHamilton_Engle.pdf