









Time Series Analysis

Chapter 5: Conditional volatility: ARCH and GARCH



Daniel Traian PELE

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

Acronyms used in this material (1/2)

- **ACF**: Autocorrelation Function
- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **AR**: AutoRegressive (model)
- **AR-GARCH**: AR model for the mean with GARCH errors
- **ARCH**: AutoRegressive Conditional Heteroskedasticity
- **ARCH-LM**: ARCH Lagrange Multiplier (test)
- **ARMA**: AutoRegressive Moving Average (model)
- **ARMA-GARCH**: ARMA model for the mean with GARCH errors
- **BET**: Bucharest Exchange Trading (index)
- **BIC**: Bayesian Information Criterion
- **BNR**: Banca Națională a României (the National Bank of Romania)
- **BVB**: Bursa de Valori București (the Bucharest Stock Exchange)
- **COVID-19**: COronaVIrus Disease 2019
- **DM**: Diebold–Mariano (test)
- **EGARCH**: Exponential GARCH
- **EWMA**: Exponentially Weighted Moving Average

Acronyms used in this material (2/2)

- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GARCH-N**: GARCH with Normal innovations
- **GJR**: Glosten–Jagannathan–Runkle (GARCH model)
- **GJR-GARCH**: Glosten–Jagannathan–Runkle GARCH
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors)
- **IGARCH**: Integrated GARCH
- **LM**: Lagrange Multiplier
- **LR**: Likelihood Ratio (test)
- **MA**: Moving Average
- **MLE**: Maximum Likelihood Estimation
- **NIC**: News Impact Curve
- **OLS**: Ordinary Least Squares
- **PACF**: Partial AutoCorrelation Function
- **QLIKE**: Quasi-LIKElihood (loss function)
- **QMLE**: Quasi-Maximum Likelihood Estimation
- **QQ**: Quantile–Quantile (plot)
- **SE**: Standard Error
- **US**: United States
- **VAR**: Vector AutoRegression

Today's question and route

- **Question**: if tomorrow's return cannot be predicted, can tomorrow's **risk** be predicted?
  - Chapters 1–3 modelled the conditional **mean** of a series; this chapter models its conditional **variance**
- **Route** of the chapter
  - stylised facts of returns; the ACF of squared returns and the ARCH-LM test
  - conditional mean and conditional variance; ARCH(q) and GARCH(1,1); persistence, IGARCH and EWMA
  - estimation by maximum likelihood with the `arch` package; Student-t innovations; ARMA-GARCH models
  - asymmetry (GJR, EGARCH), diagnostics, variance forecasts and their evaluation, VaR 1%
- Data: S&P 500, BET, EUR/RON (BNR reference rate) and Bitcoin, daily; Seminar 5 comes before this lecture

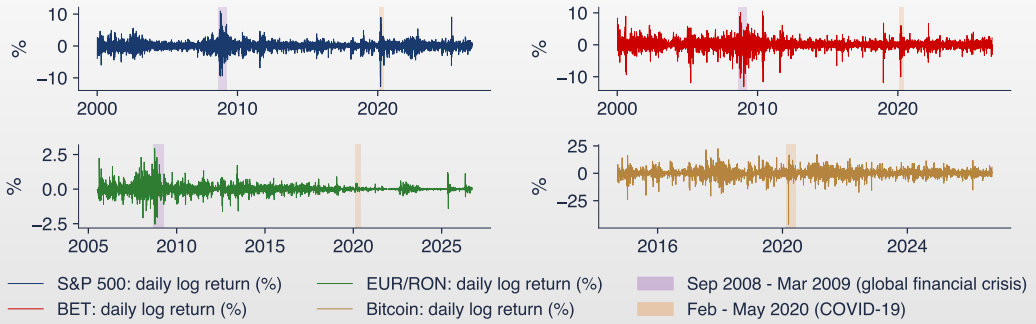
Learning outcomes

- List the stylised facts of financial returns and test for ARCH effects with the ACF of squared returns and the ARCH-LM test
- Distinguish the conditional from the unconditional variance, and explain why returns can be uncorrelated yet dependent
- Write ARCH(q), GARCH(1,1) and ARMA-GARCH models; compute the persistence, the long-run variance and the half-life
- Estimate GARCH models by maximum likelihood in Python, choose the innovation distribution and check the standardised residuals
- Measure asymmetry with GJR-GARCH, EGARCH and the news impact curve
- Forecast the variance over several horizons, evaluate the forecasts with QLIKE and compute a one-day VaR 1%

Reading and tools

- Textbook: Huang and Petukhina (2022), Ch. 6 (financial time series: ARCH and GARCH models)
  - further reading: Tsay (2010), Ch. 3; Franke, Härdle and Hafner (2019), Ch. 13
- Theory: Hamilton (1994), Ch. 21 (time series models of heteroskedasticity)
- Python Quantlets of this chapter: Quantlets/Ch_05
  - estimation with the package `arch` (install it in Colab with `pip install arch`); tests with `statsmodels`
- Lecture notebook: open in Google Colab
- Video course: Applied Time Series Analysis with Python

Daily log returns in four markets



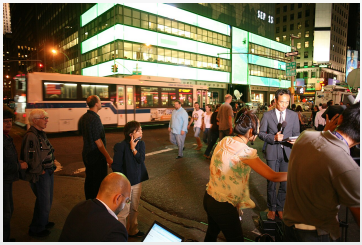
- Daily log returns $r_t = 100(\ln P_t - \ln P_{t-1})$, in %, to 18 September 2026: S&P 500 and BET since 2000, EUR/RON (BNR reference rate) since July 2005, Bitcoin since September 2014
- Shaded: the global financial crisis (September 2008 – March 2009) and the COVID-19 crash (February – May 2020)

TSA_ch5_stylised_facts

Interpreting the four series

- Calm years alternate with storms: large changes follow large changes, of either sign
  - this is **volatility clustering**, first described by Mandelbrot (1963)
- S&P 500: daily standard deviation 0.42% in 2017, 3.52% from September 2008 to March 2009 ($\times$8.3) and 3.70% in February–May 2020 ($\times$8.8)
- EUR/RON: 0.78% in the 2008 crisis, only 0.10% in 2020: the BNR kept the leu almost fixed
- A model with a constant variance (white noise, Chapter 1) describes none of these series

The crises of 2008 and 2020



Lehman Brothers headquarters, New York, 15 September 2008

Photo: Robert Scoble (2008); CC BY 2.0; Wikimedia Commons



The Financial District of New York, 25 March 2020

Photo: Billie Grace Ward (2020); CC BY 2.0; Wikimedia Commons

- 15 September 2008: Lehman Brothers files for bankruptcy; March 2020: the COVID-19 pandemic closes economies
- In both episodes calm returned only after months: a shock to volatility is persistent

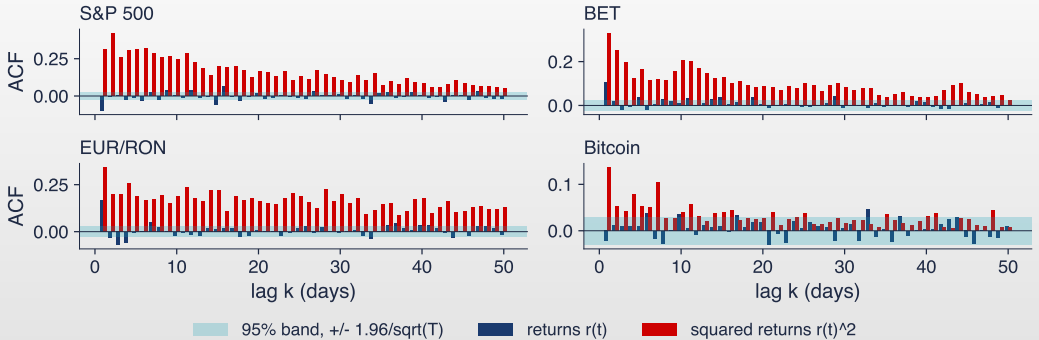
Stylised facts in numbers

	T	s.d.	skew.	kurt.	$\hat{\rho}_1(r)$	$\hat{\rho}_1(r^2)$	$Q(10)$, r	$Q(10)$, r^2	LM(5)
S&P 500	6 714	1.21	−0.35	13.7	−0.10	0.31	96	6133	1684
BET	6 670	1.40	−0.65	14.1	0.11	0.33	108	2440	982
EUR/RON	5 352	0.28	0.73	18.1	0.17	0.34	222	2420	871
Bitcoin	4 384	3.49	−0.70	15.0	−0.02	0.14	20	213	115

- **Heavy tails**: kurtosis between 13.7 and 18.1, far above 3, the value of the Normal distribution
  - kurtosis $K = E[(r-\mu)^4]/\sigma^4$; skewness $E[(r-\mu)^3]/\sigma^3$ (Chapter 1)
- **Little autocorrelation in r_t**, strong autocorrelation in r_t^2: $Q(10)$ of the squares is 11–64 times larger
- $Q(10)$: the statistic of Ljung and Box (1978) (Chapter 1), $\chi^2(10)$, 5% value 18.31; LM(5): the ARCH-LM test (two slides ahead), $\chi^2(5)$, 5% value 11.07

TSA_ch5_stylised_facts

Autocorrelation of returns and of squared returns



- Sample ACF (autocorrelation function, Chapter 1) of r_t and of r_t^2, lags 1–50, with the band $\pm 1.96/\sqrt{T}$ of an i.i.d. (independent and identically distributed) series

TSA_ch5_stylised_facts

Interpreting the two ACFs

- Returns: almost all autocorrelations inside the band (BET: $\hat{\rho}_1 = 0.11$, a thin market; Chapter 2)
  - the direction of tomorrow's move is almost unpredictable
- Squared returns: S&P 500 0.31 at lag 1 and still 0.05 at lag 50; 50 of 50 lags outside the band
  - the size of tomorrow's move is predictable: a large move today makes a large move tomorrow more likely
- Uncorrelated is not the same as independent: r_t is close to white noise but not i.i.d. (Chapter 1)
- Bitcoin: weaker but significant ($\hat{\rho}_1(r_t^2) = 0.14$): a few huge days dominate the squares

Testing for ARCH effects

- **ARCH effects**: the conditional variance depends on the past (the name comes from the ARCH model, Section 3)
- **Ljung–Box on squares** (the test of McLeod and Li (1983)): $Q(m) = T(T+2)\sum_{k=1}^{m} \hat{\rho}_k^2(\hat{\varepsilon}^2)/(T-k)$
  - $\hat{\rho}_k(\hat{\varepsilon}^2)$: the sample autocorrelation of the squared residuals at lag k; m: the number of lags; T: the number of observations
  - $\hat{\varepsilon}_t$: residuals of the mean model (for returns, $r_t - \bar{r}$); under H_0 (no ARCH effects), $Q(m) \sim \chi^2(m)$
- **ARCH-LM test** Engle (1982): LM (Lagrange multiplier) test from the auxiliary regression
  - $\hat{\varepsilon}_t^2 = b_0 + b_1\hat{\varepsilon}_{t-1}^2 + \cdots + b_q\hat{\varepsilon}_{t-q}^2 + u_t$, estimated by OLS (ordinary least squares)
  - $b_0, \ldots, b_q$: the coefficients; q: the number of lagged squares; R^2: the coefficient of determination of this regression
  - H_0: $b_1 = \cdots = b_q = 0$; LM $= nR^2 \sim \chi^2(q)$, $n =$ number of observations in the regression
- Both tests are run on the residuals of the mean model, before and after fitting a volatility model

Worked example: ARCH-LM on the BET

- Mean model: AR(1) for the BET returns (Chapter 2), $\hat{\phi} = 0.109$; residuals $\hat{\varepsilon}_t$; $q = 5$
- Auxiliary regression (OLS)
  - $\hat{\varepsilon}_t^2 = 0.862 + 0.276\,\hat{\varepsilon}_{t-1}^2 + 0.099\,\hat{\varepsilon}_{t-2}^2 + 0.088\,\hat{\varepsilon}_{t-3}^2 + (-0.015)\,\hat{\varepsilon}_{t-4}^2 + 0.102\,\hat{\varepsilon}_{t-5}^2$
  - $R^2 = 0.1526$, $n = 6\,664$
- Step by step: LM $= nR^2 = 6\,664 \times 0.1526 = 1017$; 5% critical value $\chi^2_{0.95}(5) = 11.07$
  - $1017 \gg 11.07$: reject H_0; the variance of the BET depends on the last five squared shocks
- Interpretation: $R^2 = 0.1526$ is small, but for squared returns it is a strong signal; the model of the variance is the next step

TSA_ch5_mean_model

Recap: Stylised facts

- Heavy tails, volatility clustering, little autocorrelation in r_t, strong autocorrelation in r_t^2
- ARCH effects: Ljung–Box on squared residuals, ARCH-LM $= nR^2 \sim \chi^2(q)$
- All four series reject H_0 by a wide margin

Conditioning on the past

- $\mathcal{F}_{t-1}$: the information available at the end of day $t-1$ (all past values)
  - conditional expectation $E[X \mid \mathcal{F}_{t-1}]$: the best forecast of X given the past (Chapter 2)
- **Conditional mean**: $\mu_t = E[r_t \mid \mathcal{F}_{t-1}]$
  - an ARMA model (Chapter 2) is a model for μ_t; for daily returns μ_t is small and almost constant
- **Conditional variance**: $\sigma_t^2 = \mathrm{Var}(r_t \mid \mathcal{F}_{t-1}) = E[(r_t - \mu_t)^2 \mid \mathcal{F}_{t-1}]$
  - σ_t is the **conditional volatility**: known at $t-1$, it changes from day to day
  - **conditional heteroskedasticity**: a conditional variance that is not constant
- **Unconditional variance** $\mathrm{Var}(r_t)$: one number for the whole sample; it can be constant while σ_t^2 moves

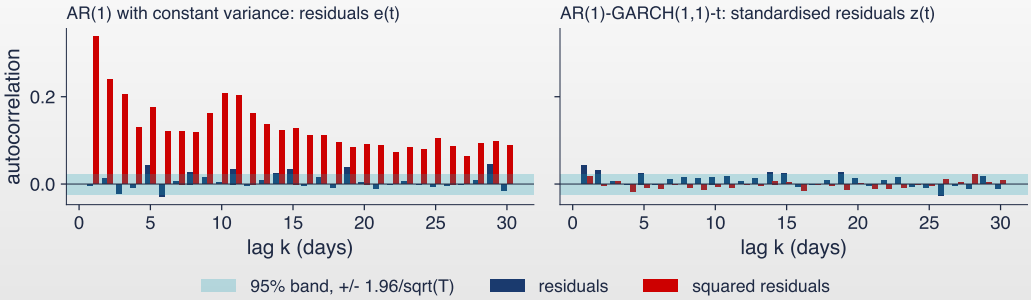
The basic decomposition

- Every model of this chapter: $r_t = \mu_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t$
  - z_t: the **innovations**, i.i.d. with mean 0 and variance 1 (Normal, Student-t, ...)
  - ε_t: the **shock**; μ_t and σ_t depend only on the past
- Consequences
  - $E[\varepsilon_t \mid \mathcal{F}_{t-1}] = \sigma_t E[z_t] = 0$: the shocks are a **martingale difference**, hence uncorrelated (white noise)
  - $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$: the variance is predictable
  - the ε_t are uncorrelated but not independent: ε_t^2 is correlated with ε_{t-1}^2
- Two models, one for each moment: ARMA for μ_t (Chapter 2), GARCH for σ_t^2 (this chapter); together: **ARMA-GARCH**

Worked example: conditional and unconditional variance

- A market has calm days ($\sigma_t^2 = 0.5$) and stormy days ($\sigma_t^2 = 4.5$), each with probability $1/2$; the mean is 0
  - law of total variance: $\mathrm{Var}(r_t) = E[\mathrm{Var}(r_t \mid \mathcal{F}_{t-1})] + \mathrm{Var}(E[r_t \mid \mathcal{F}_{t-1}])$
  - step by step: $0.5 \times 0.5 + 0.5 \times 4.5 + 0 = 2.50$; unconditional standard deviation 1.58%
- The unconditional variance is an average; the conditional variance tells which day we are in
  - a calm day: $\sigma_t = 0.71\%$; a stormy day: $\sigma_t = 2.12\%$; the average 1.58% fits neither
- With Normal z_t, the mixture has kurtosis $3E[\sigma_t^4]/(E[\sigma_t^2])^2 = 4.92$: a time-varying variance creates heavy tails

The mean model leaves ARCH effects



- BET: left, an AR(1) for the mean (order chosen by BIC among AR(0)–AR(5): $p = 1$), constant variance; right, the same AR(1) with a GARCH(1,1)-t variance (Section 6)

TSA_ch5_mean_model

Interpreting the residual ACFs

- Left: the residuals look almost like white noise, the Box–Jenkins check of Chapter 2 would be nearly passed
  - but their squares: $\hat{\rho}_1 = 0.34$, $Q(10) = 2496$: strong ARCH effects
- Right: after dividing by $\hat{\sigma}_t$, the squares are clean: $\hat{\rho}_1 = 0.018$, $Q(10) = 7.2$ (p $= 0.705$), ARCH-LM 4.8 (p $=$ 0.446)
  - some autocorrelation in the level remains ($Q(10) = 28.6$): the BET mean has more structure than an AR(1)
- Lesson: check both $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_t^2$; a white-noise check on $\hat{\varepsilon}_t$ alone misses the variance

Recap: Conditional mean and variance

- $r_t = \mu_t + \sigma_t z_t$: μ_t from an ARMA model, σ_t^2 from a volatility model
- Shocks are uncorrelated but dependent through their squares
- Unconditional variance = average of the conditional variances; the mixture has heavy tails

1982: Robert Engle and ARCH

- Engle (1982): **ARCH** (autoregressive conditional heteroskedasticity): today's variance depends on yesterday's squared shocks
  - first application: the variance of inflation in the United Kingdom, not a financial market
- Nobel Prize 2003: "for methods of analyzing economic time series with time-varying volatility (ARCH)"
  - shared with Clive Granger (Chapter 3: spurious regression; Chapter 7: cointegration); Nobel lecture: Engle (2004)



Robert F. Engle (2022)

Photo: Jon Demske (2022); CC BY-SA 4.0; Wikimedia Commons



Stockholm Concert Hall, venue of the Nobel Prize ceremony

Photo: Karen Zhou (2025); CC BY-SA 4.0; Wikimedia Commons

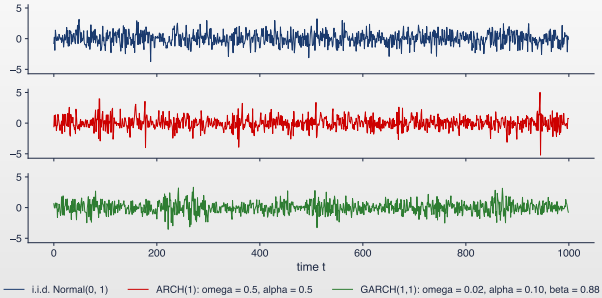
ARCH(1): definition

- $\varepsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2, \quad \omega > 0,\ \alpha \geq 0$
  - ω: the minimum level of the variance; α: how strongly the variance reacts to yesterday's squared shock
  - $\omega > 0$ and $\alpha \geq 0$ keep the variance positive
  - a large shock yesterday, of either sign, raises today's variance
- "Autoregressive": ε_t^2 follows an AR(1) model (Chapter 2)
  - write $\varepsilon_t^2 = \sigma_t^2 + v_t$, with $v_t = \sigma_t^2(z_t^2 - 1)$ and $E[v_t \mid \mathcal{F}_{t-1}] = 0$
  - then $\varepsilon_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + v_t$: an AR(1) for the squared shocks, with white-noise errors v_t
- Hence the ACF of ε_t^2 is $\rho_k = \alpha^k$ (when the fourth moment exists): clustering, but a fast decay

ARCH(1): properties

- $E[\varepsilon_t] = 0$ and $\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$ for $k \neq 0$: white noise
- **Unconditional variance** (stationary if $\alpha < 1$): take expectations in $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2$
  - $\bar{\sigma}^2 = E[\varepsilon_t^2] = \omega + \alpha\bar{\sigma}^2$, so $\bar{\sigma}^2 = \omega/(1-\alpha)$
- **Kurtosis** with Normal z_t (if $3\alpha^2 < 1$): $K = 3\,\dfrac{1-\alpha^2}{1-3\alpha^2} > 3$
  - Normal innovations, yet heavy-tailed shocks; if $\alpha \geq 1/\sqrt{3} = 0.577$, the fourth moment is infinite
- **Worked example**: $\omega = 0.5$, $\alpha = 0.5$
  - $\bar{\sigma}^2 = 0.5/0.5 = 1.00$; $K = 3 \times 0.75/0.25 = 9.0$
  - after a shock $\varepsilon_{t-1} = 2$: $\sigma_t^2 = 0.5 + 0.5 \times 4 = 2.50$, $\sigma_t = 1.58$: the variance more than doubles

Simulated paths: i.i.d., ARCH(1) and GARCH(1,1)



- Three series of 1000 values with unconditional variance 1: i.i.d. Normal; ARCH(1) with $\omega = \alpha = 0.5$; GARCH(1,1) with $\omega = 0.02$, $\alpha = 0.10$, $\beta = 0.88$ (next section)
- Sample kurtosis: i.i.d. 3.12, ARCH 4.91, GARCH 3.74 (theory: 3, 9.0, 6.06): the fourth moment converges slowly
- Interpretation: ARCH(1) gives isolated spikes; GARCH(1,1) gives long calm and long agitated periods, as in real returns

TSA_ch5_simulation_likelihood

ARCH(q) and its limits

- **ARCH(q)**: $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$, $\alpha_i \geq 0$; stationary if $\sum_i \alpha_i < 1$
  - a shock affects the variance for exactly q days; real clustering lasts for months

S&P 500, Normal z_t	parameters	log-likelihood	AIC	BIC	$\sum \alpha_i$ $(+\beta)$
ARCH(1)	3	−10305.5	20617.0	20637.5	0.394
ARCH(5)	7	−9340.5	18695.1	18742.7	0.816
ARCH(10)	12	−9227.6	18479.3	18561.0	0.849
GARCH(1,1)	4	−9218.0	**18444.0**	**18471.2**	0.980

- AIC (Akaike) and BIC (Bayesian information criterion): AIC $= -2\ell + 2k$, BIC $= -2\ell + k \ln T$ (Chapter 2), ℓ: the maximised log-likelihood, k: parameters; smaller is better
- Interpretation: ARCH needs ten lags and still loses to GARCH(1,1), which has four parameters: the motivation for GARCH

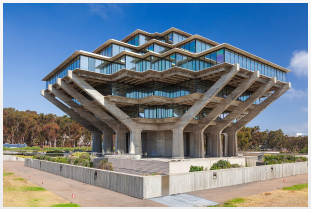
TSA_ch5_garch_estimation

Recap: The ARCH model

- $\sigma_t^2 = \omega + \sum_i \alpha_i \varepsilon_{t-i}^2$: an AR model for the squared shocks
- Unconditional variance $\omega/(1 - \sum \alpha_i)$; kurtosis above 3 even with Normal innovations
- Real clustering needs many lags: GARCH replaces them with one extra parameter

1986: Tim Bollerslev and GARCH

- Bollerslev (1986), a doctoral student of Engle at the University of California, San Diego: add yesterday's variance to the equation
  - **GARCH** (generalised ARCH): a few parameters replace a long ARCH(q)
- Student-t innovations for returns: Bollerslev (1987)
- Still the benchmark: Hansen and Lunde (2005) compared 330 volatility models and found GARCH(1,1) hard to beat on exchange rates



Geisel Library, University of California, San Diego

Photo: Stephen Bay (2010); CC BY 4.0; Wikimedia Commons

GARCH(1,1): definition

- $\varepsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \omega + \alpha\,\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2, \quad \omega > 0,\ \alpha \geq 0,\ \beta \geq 0$
  - GARCH(p,q) adds more lags of both kinds; in practice (1,1) is almost always enough
- Reading the parameters
  - α: the **reaction** to yesterday's news; β: the **memory** of the variance
  - $\alpha + \beta$: the **persistence**; ω sets the long-run level
- By recursion: an ARCH(∞) with geometric weights
  - $\sigma_t^2 = \dfrac{\omega}{1-\beta} + \alpha \sum_{j=0}^{\infty} \beta^j\, \varepsilon_{t-1-j}^2$: all past shocks matter, with weights that decay like β^j

GARCH(1,1) as an ARMA(1,1) for the squared shocks

- As for ARCH, write $\varepsilon_t^2 = \sigma_t^2 + v_t$ and substitute $\sigma_t^2 = \varepsilon_t^2 - v_t$ in the equation
  - $\varepsilon_t^2 = \omega + (\alpha + \beta)\,\varepsilon_{t-1}^2 + v_t - \beta\, v_{t-1}$
  - an ARMA(1,1) (Chapter 2) with AR coefficient $\alpha + \beta$ and MA coefficient $-\beta$
- Consequences
  - **covariance stationary** if $\alpha + \beta < 1$ (the AR root outside the unit circle)
  - the ACF of ε_t^2 decays like $(\alpha + \beta)^k$: slow when the persistence is close to 1, as in the ACF of real squared returns
  - identification by the ACF and the PACF (partial autocorrelation function) of ε_t^2 is possible in principle, but in practice we use maximum likelihood (Section 5)

Long-run variance, kurtosis, persistence and half-life

- **Long-run (unconditional) variance**: $\bar{\sigma}^2 = \omega/(1-\alpha-\beta)$; annualised volatility $\sqrt{252\,\bar{\sigma}^2}$ for 252 trading days
  - kurtosis with Normal z_t: $K = 3\,\dfrac{1-(\alpha+\beta)^2}{1-(\alpha+\beta)^2-2\alpha^2} > 3$; $\alpha = 0.10$, $\beta = 0.88$: $K = 6.06$
- A deviation from $\bar{\sigma}^2$ shrinks by the factor $\alpha+\beta$ each day (Section 9): $E_t[\sigma^2_{t+h}] - \bar{\sigma}^2 = (\alpha+\beta)^{h-1}(\sigma^2_{t+1} - \bar{\sigma}^2)$
  - $E_t[\cdot]$: the expectation given the information of day t; σ^2_{t+1}: tomorrow's variance, already known at t
  - **half-life**: the number of days after which half of a variance shock is gone, $h_{1/2} = \ln 0.5/\ln(\alpha+\beta)$
  - $\alpha+\beta = 0.90$: 6.6 days; 0.98: 34.3 days; 0.99: 69.0 days: small changes near 1 matter a lot

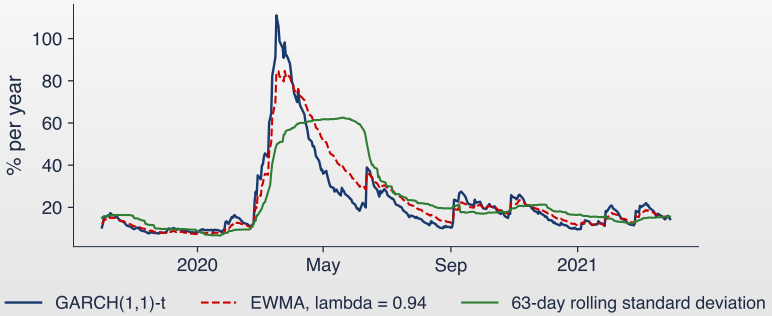
Worked example: one GARCH(1,1) step

- Daily returns in %: $\omega = 0.02$, $\alpha = 0.10$, $\beta = 0.88$, $\mu = 0$
- Long-run quantities
  - persistence $\alpha + \beta = 0.98$; $\bar{\sigma}^2 = 0.02/(1 - 0.98) = 1.00$; annualised volatility $\sqrt{252 \times 1.00} = 15.9\%$
  - half-life $\ln 0.5 / \ln 0.98 = 34.3$ days
- Today: $\sigma_t^2 = 1.2$ and the return is $r_t = -3\%$
  - tomorrow: $\sigma_{t+1}^2 = 0.02 + 0.10 \times 9 + 0.88 \times 1.2 = 1.976$, $\sigma_{t+1} = 1.41\%$
  - the shock raised the variance from 1.2 to almost 2; it then decays towards 1.00 with the factor 0.98 per day

IGARCH and the link with EWMA

- **IGARCH** (integrated GARCH) Engle and Bollerslev (1986): $\alpha + \beta = 1$: a unit root in the ARMA(1,1) for ε_t^2 (Chapter 3)
  - shocks to the variance never die out: no half-life, no finite unconditional variance
- **EWMA** (exponentially weighted moving average) of J.P. Morgan/Reuters (1996): $\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2$
  - $\lambda \in (0,1)$: the weight of yesterday's variance; $1-\lambda$: the weight of yesterday's squared return
  - simple exponential smoothing (Chapter 0) applied to r_t^2, with smoothing constant $1-\lambda$
  - an IGARCH with $\omega = 0$, $\alpha = 1-\lambda$, $\beta = \lambda$; RiskMetrics uses $\lambda = 0.94$ for daily data (half-life of the weights 11.2 days)
- **Question for the room**: right after a crash, an analyst forecasts the volatility one year ahead with EWMA. What goes wrong?
  - **Answer**: EWMA keeps today's crisis level for the whole year; GARCH lets it decay towards the long-run level

GARCH, EWMA and a rolling window in 2020



- S&P 500, annualised volatility: GARCH(1,1)-t estimated on all data, EWMA with $\lambda = 0.94$, and the standard deviation of the last 63 days (Chapter 0: a moving average)

TSA_ch5_volatility_markets

Interpreting the three estimates

- Peaks: GARCH 111% (17 March 2020), EWMA 85% (25 March 2020), 63-day window 63% (20 May 2020)
- The rolling window reacts late and falls abruptly when the crash leaves the window (the "ghost" of the crash)
  - on 30 June 2020: GARCH 28%, EWMA 30%, window 32%
- GARCH reacts fastest (α larger than $1 - \lambda$) and decays fastest (mean reversion); EWMA lies in between

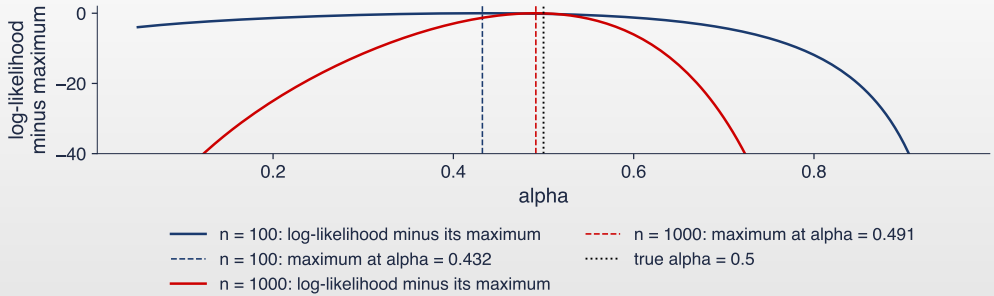
Recap: The GARCH(1,1) model

- $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$: reaction α, memory β; an ARMA(1,1) for ε_t^2
- $\alpha + \beta < 1$: long-run variance $\omega/(1 - \alpha - \beta)$, half-life $\ln 0.5 / \ln(\alpha + \beta)$
- $\alpha + \beta = 1$: IGARCH; EWMA is an IGARCH without ω

The likelihood of a GARCH model

- The returns are dependent, so the joint density is a product of conditional densities
  - $f(r_1, \ldots, r_T) = f(r_1) \prod_{t=2}^{T} f(r_t \mid \mathcal{F}_{t-1})$, with $r_t \mid \mathcal{F}_{t-1} \sim N(\mu, \sigma_t^2)$ for Normal innovations
- **Conditional log-likelihood**, $\theta = (\mu, \omega, \alpha, \beta)$:
  - $\ell(\theta) = -\frac{1}{2} \sum_{t=2}^{T} \left[\ln(2\pi) + \ln \sigma_t^2(\theta) + \frac{(r_t - \mu)^2}{\sigma_t^2(\theta)} \right]$
  - $\sigma_t^2(\theta)$ is computed recursively from a starting value σ_1^2 (for example the sample variance)
- **MLE** (maximum likelihood estimation): $\hat{\theta} = \arg\max_\theta \ell(\theta)$, under $\omega > 0$, $\alpha, \beta \geq 0$, $\alpha + \beta < 1$
  - no closed form: a numerical optimiser searches for the maximum; OLS is not possible, since σ_t^2 is not observed

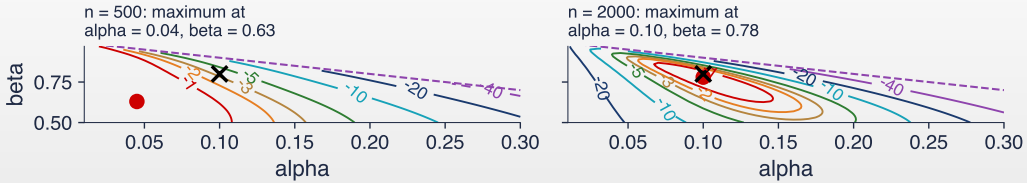
The likelihood of ARCH(1)



- Simulated ARCH(1) with $\omega = \alpha = 0.5$; $\ell(\alpha)$ with $\omega = 1 - \alpha$; each curve minus its maximum
- $n = 100$: $\hat{\alpha} = 0.432$ (standard error 0.127); $n = 1000$: $\hat{\alpha} = 0.491$ (0.035)
- Interpretation: more data make the curve sharper; its curvature at the maximum gives the standard error

TSA_ch5_simulation_likelihood

The likelihood of GARCH(1,1)



- Simulated GARCH(1,1) with $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.8$; ω set so that the unconditional variance equals the sample variance; contours: log-likelihood minus its maximum
- $n = 500$: maximum at $(0.045, 0.63)$, far from the truth; $n = 2000$: $(0.100, 0.78)$
- Interpretation: a long, flat ridge along which α and β trade off: GARCH needs long samples

TSA_ch5_simulation_likelihood

Estimation step by step and with the `arch` package

S&P 500, GARCH(1,1)-N	step by step (SE)	`arch`	classic SE	robust SE
μ	0.0614 (0.0098)	0.0615	0.0098	0.0100
ω	0.0253 (0.0030)	0.0254	0.0030	0.0048
α	0.1205 (0.0086)	0.1206	0.0086	0.0118
β	0.8600 (0.0091)	0.8599	0.0091	0.0125

- Step by step: $-\ell(\theta)$ written in Python and minimised with `scipy.optimize.minimize` under $\alpha + \beta < 1$
  - with the package: `arch_model(r, mean='Constant', vol='GARCH', p=1, q=1, dist='normal').fit()`
- SE (standard error), classic: from the inverse Hessian (the curvature of ℓ); robust: valid when z_t is not Normal Bollerslev and Wooldridge (1992)
- Interpretation: same estimates by both routes; the robust SE are about 1.4 times larger: with heavy-tailed z_t the classic SE overstate the precision

TSA_ch5_garch_estimation

Quasi-maximum likelihood and practical rules

- **QMLE** (quasi-maximum likelihood estimation): maximise the Normal likelihood even if z_t is not Normal
  - the estimates stay consistent if μ_t and σ_t^2 are correctly specified, but only the robust ("sandwich") SE are valid Bollerslev and Wooldridge (1992)
- Practical rules
  - returns in % (not decimals): the optimiser fails when ω is tiny; `arch` warns about badly scaled data
  - EUR/RON moves about 0.3% a day: we estimate on $10\,r_t$ and convert ω back (divide by 100)
  - use at least 1000–2000 daily observations; check the convergence flag of the optimiser
  - an estimate on the boundary ($\hat{\alpha} + \hat{\beta} = 1$ or $\hat{\alpha} = 0$) has no usual standard error: report it as such

Student-t and skewed-t innovations

- GARCH with Normal z_t explains only part of the kurtosis: S&P 500 returns 13.7, standardised residuals $\hat{z}_t = \hat{\varepsilon}_t/\hat{\sigma}_t$ still 4.7
  - the remaining heavy tails must come from the distribution of z_t
- **Standardised Student-t** with $\nu > 2$ degrees of freedom Bollerslev (1987)
  - $z_t = t_\nu \sqrt{(\nu-2)/\nu}$ has variance 1; kurtosis $3 + 6/(\nu-4)$ for $\nu > 4$; $\nu \to \infty$: the Normal distribution
  - ν is estimated together with $(\mu, \omega, \alpha, \beta)$: `dist='t'` in `arch`
- **Skewed t** Hansen (1994): one more parameter $\lambda \in (-1, 1)$; $\lambda < 0$: a longer left tail (`dist='skewt'`)
  - nested models: Normal $\subset$ t $\subset$ skewed t; compare them with LR (likelihood ratio) tests or AIC/BIC

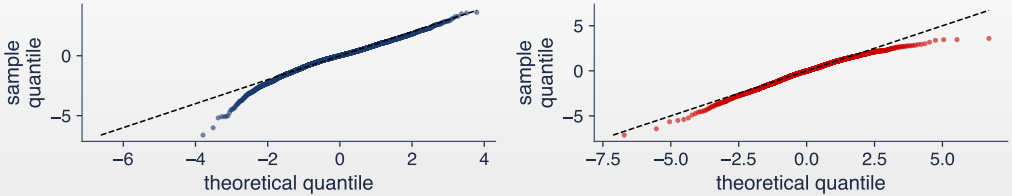
S&P 500: three innovation distributions

GARCH(1,1)	ω	α	β	ν	λ	log-lik.	$\alpha+\beta$
Normal	0.0254	0.121	0.860	–	–	−9218.0	0.980
Student-t	0.0157	0.123	0.872	6.26	–	−9062.5	0.995
skewed t	0.0153	0.120	0.872	6.92	−0.114	−9038.4	0.993

- LR statistic $2(\ell_1 - \ell_0)$, ℓ_1 and ℓ_0: the log-likelihoods of the larger and of the smaller model: t against Normal 311.0 (1 restriction, $\chi^2_{0.95}(1) = 3.84$); skewed t against t 48.1: both reject the smaller model
- $\hat{\nu} = 6.26$: tails far heavier than Normal; $\hat{\lambda} = -0.114$: large falls more frequent than large rises
- Interpretation: α and β hardly change; the innovation distribution matters for tail quantiles (VaR), less for σ_t

TSA_ch5_garch_estimation

QQ plots of the standardised residuals



- QQ (quantile–quantile) plot: sample quantiles of $\hat{z}_t$ against the quantiles of the assumed distribution; a straight line means a good fit
- Left, GARCH-N against the Normal: 0.1% quantile -4.81, against -3.09 in theory; right, GARCH-t against the standardised t ($\hat{\nu} = 6.3$): -4.79 against -4.19
- Interpretation: the t fits much better, but the largest falls are still deeper: the left tail is the problem (asymmetry, Section 7)

TSA_ch5_garch_estimation

Recap: Estimation by maximum likelihood

- $\ell(\theta)$ = sum of conditional log-densities; $\sigma_t^2(\theta)$ computed recursively; numerical maximisation
- The likelihood has a ridge in (α, β): long samples; report robust SE
- Student-t or skewed-t innovations for the remaining heavy tails

GARCH(1,1)-t estimates

	from	ω	α	β	ν	$\alpha+\beta$	half-life	long-run vol.	sample vol.
S&P 500	2000	0.0157	0.123	0.872	6.26	0.9947	131	27.3	19.2
BET	2000	0.0448	0.192	0.799	5.01	0.9907	74	34.7	22.1
EUR/RON	2005	0.0000	0.123	0.877	4.20	1.000	–	–	4.4
Bitcoin	2014	0.1937	0.109	0.891	3.18	1.000	–	–	66.6

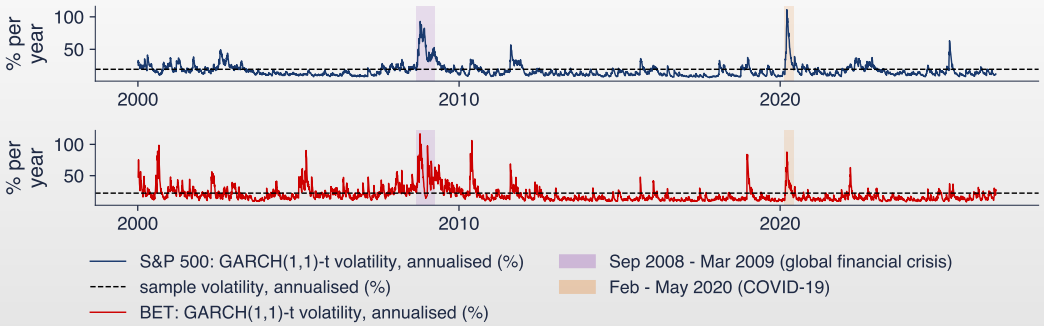
- Daily log returns in % to 18 September 2026; half-life in days; volatilities annualised with the actual number of observations per year (Bitcoin: 365), in %
- S&P 500 and BET: $\alpha+\beta$ 0.9947 and 0.9907, half-lives of 131 and 74 trading days; BET reacts more to news ($\alpha = 0.192$)
- EUR/RON and Bitcoin: $\hat{\alpha}+\hat{\beta}=1$ on the boundary: IGARCH; every $\hat{\nu}$ between 3.18 and 6.26: heavy tails everywhere

TSA_ch5_volatility_markets

Interpreting IGARCH and the long-run volatility

- EUR/RON: $\hat{\omega} \approx 0$ and $\hat{\beta} = 0.88$: an EWMA with $\lambda \approx 0.88$; Bitcoin: $\hat{\beta} = 0.89$
  - no half-life and no long-run volatility: the volatility level drifts like a random walk (Chapter 3)
  - EUR/RON: a managed rate whose volatility fell from one regime to the next (the BNR policy changes over time)
- **Question for the room**: S&P 500: long-run volatility 27.3%, sample volatility 19.2%. Is the model wrong?
  - **Answer**: not necessarily: $\bar{\sigma}^2 = \omega/(1 - \alpha - \beta)$ divides by $1 - \alpha - \beta = 0.0053$, a small and imprecise number
  - the Normal GARCH gives 18.1%: the long-run level is the least reliable output of a persistent GARCH

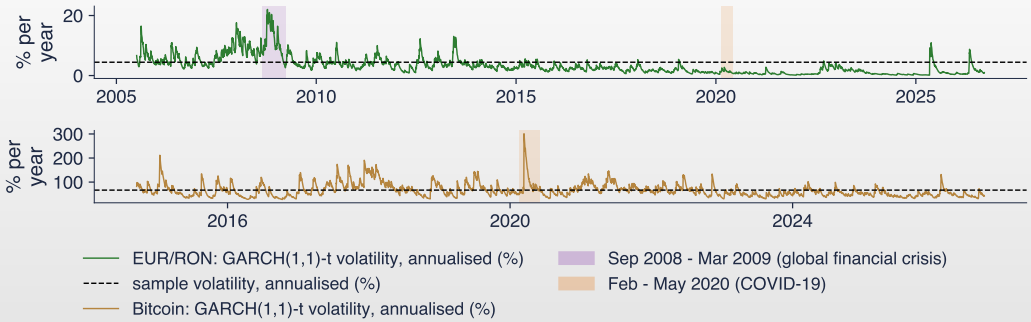
Conditional volatility: S&P 500 and BET



- GARCH(1,1)-t volatility $\hat{\sigma}_t$, annualised; dashed: the sample volatility
- S&P 500: peaks 93% (16 October 2008) and 111% (17 March 2020); BET: 117% (15 October 2008) and 88% (18 March 2020); medians 14.3% and 15.6%

TSA_ch5_volatility_markets

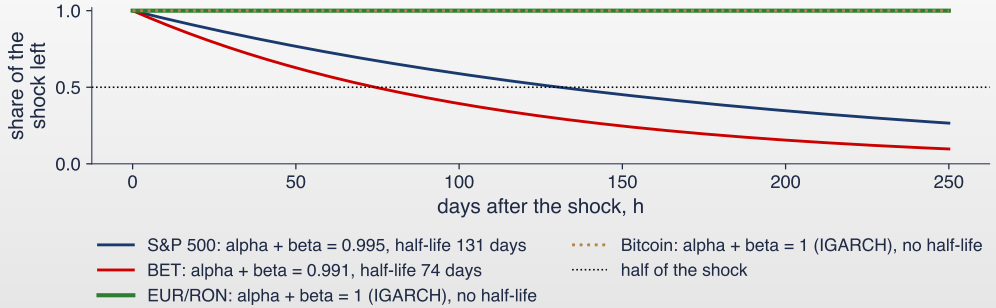
Conditional volatility: EUR/RON and Bitcoin



- EUR/RON: 22% at the peak of the 2008 crisis (13 October 2008), only 3% in 2020; on 18 September 2026: 1.0%
- Bitcoin: 301% on 13 March 2020; median 58.7%, 4.1 times the median of the S&P 500
- Interpretation: the two IGARCH series have volatility levels that wander: no single "normal" level

TSA_ch5_volatility_markets

How long does a volatility shock last?



- Share of a variance shock left after h days: $(\alpha + \beta)^h$; the half-life is where the curve crosses 1/2
- Interpretation: after a crisis the S&P 500 needs about half a year to halve the excess variance, the BET about three months; EUR/RON and Bitcoin never

TSA_ch5_volatility_markets

ARMA-GARCH: one model for the mean and the variance

- **ARMA(p,q)-GARCH(1,1)**: $r_t = c + \sum_{i=1}^{p} \phi_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \quad \varepsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$
  - the ARMA part is the model of Chapter 2; its errors are no longer i.i.d., they follow a GARCH
- Estimation: both parts at once, by maximum likelihood
  - in `arch`: `arch_model(r, mean='AR', lags=1, vol='GARCH', dist='t')` (AR means only; for MA terms use two steps or another package)
- Why not OLS for the mean and then GARCH?
  - the OLS coefficients are still consistent, but their usual SE assume a constant variance: the t tests of Chapter 2 are wrong
  - the joint maximum likelihood estimate gives less weight to stormy days, hence more precise $\hat{\phi}$ and valid intervals

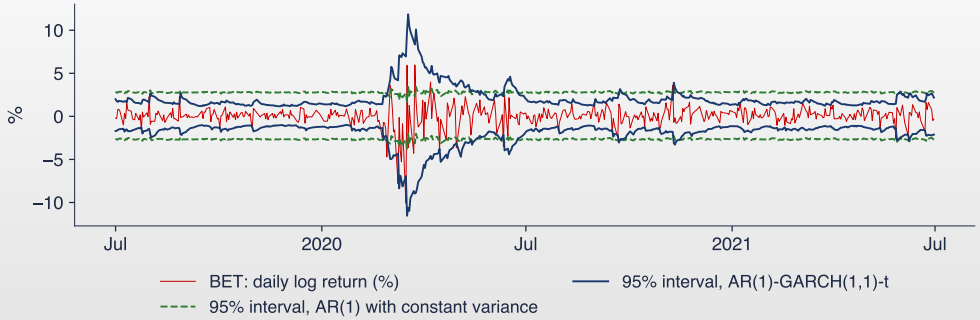
AR(1)-GARCH(1,1)-t on four series

	p (BIC)	$\hat{\phi}$, OLS (SE)	$\hat{\phi}$, AR-GARCH (SE)	$Q(10)$ of $\hat{z}_t$, const. mean	$Q(10)$ of $\hat{z}_t$, AR(1)	ΔBIC
S&P 500	1	−0.099 (0.012)	−0.054 (0.012)	21.9 (0.016)	15.1 (0.087)	−19.5
BET	1	0.109 (0.012)	0.091 (0.013)	114.6 (<0.001)	28.6 (<0.001)	−46.9
EUR/RON	3	0.170 (0.013)	0.031 (0.016)	10.5 (0.402)	8.7 (0.469)	4.2
Bitcoin	0	−0.022 (0.015)	−0.043 (0.014)	29.9 (<0.001)	48.4 (<0.001)	−9.5

- p: AR order chosen by BIC with a constant variance; SE of the AR-GARCH: robust; ΔBIC: AR(1)-GARCH-t minus GARCH-t with a constant mean (negative: AR(1) preferred)
- S&P 500: $\hat{\phi}$ shrinks from -0.099 to -0.054: the OLS value is driven by the stormy days of 2008 and 2020
- BET: a positive autocorrelation survives ($\hat{\phi} = 0.091$, $t = 6.8$): slow price adjustment in a less liquid market; EUR/RON: from 0.170 to 0.031, no longer significant
- Interpretation: the mean model and the variance model must be chosen together; the conclusions about ϕ can change

TSA_ch5_arma_garch

Forecast intervals with time-varying width



- BET, one-day-ahead 95% intervals: AR(1) with constant variance (width 5.4 points every day) and AR(1)-GARCH(1,1)-t (width from 2.2 to 22.5 points in this window)

TSA_ch5_arma_garch

Interpreting the two intervals

- Over the whole sample both miss about 5% of the days: 5.0% and 5.2%
  - the constant interval is right on average, but wrong almost every day
- In 2017 (calm): the constant interval misses 0.4% of the days, the GARCH interval 2.0%
  - in March–April 2020: 33.3% against 9.5%: the misses of the constant interval cluster in crises
- Chapter 2 gave intervals whose width depends only on the horizon; with GARCH it also depends on the current state

Recap: Four markets and ARMA-GARCH

- $\alpha + \beta$ close to 1 everywhere; EUR/RON and Bitcoin are IGARCH
- The long-run volatility is imprecise when $\alpha + \beta$ is near 1
- ARMA-GARCH: the mean of Chapter 2 with GARCH errors; joint MLE; intervals that follow the volatility

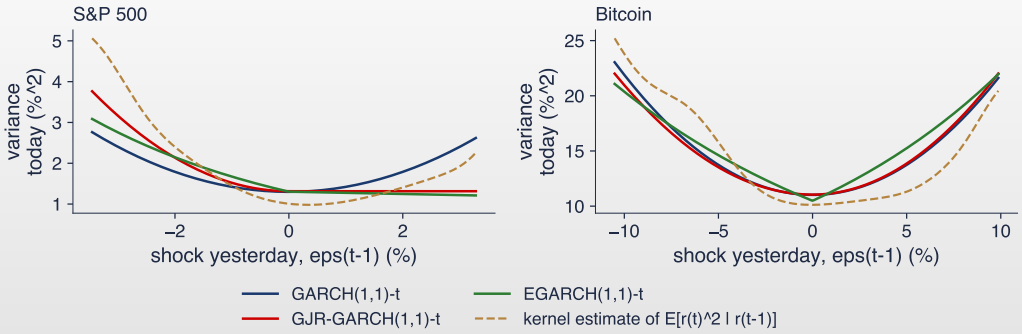
The leverage effect and the news impact curve

- **Leverage effect**: falls in stock prices raise volatility more than rises of the same size Christie (1982)
  - a lower share price raises the debt-to-equity ratio, so the equity becomes riskier; news of higher risk also lowers prices at once
- GARCH(1,1) cannot see it: σ_t^2 depends on ε_{t-1}^2, not on the sign of ε_{t-1}
  - the QQ plot and the skewed t ($\hat{\lambda} < 0$) already pointed to the left tail
- **NIC** (news impact curve) Engle and Ng (1993): σ_t^2 as a function of ε_{t-1}, with σ_{t-1}^2 fixed at the unconditional variance; for GARCH, a symmetric parabola

GJR-GARCH and EGARCH

- **GJR-GARCH** Glosten, Jagannathan and Runkle (1993): $\sigma_t^2 = \omega + (\alpha + \gamma\, I_{t-1})\, \varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, with $I_{t-1} = 1$ if $\varepsilon_{t-1} < 0$, else 0
  - good news: slope α; bad news: $\alpha + \gamma$; leverage effect: $\gamma > 0$; in `arch`: `o=1`
  - persistence $\alpha + \beta + \gamma/2$ for symmetric z_t (half of the shocks are negative)
- **EGARCH** (exponential GARCH) Nelson (1991): $\ln \sigma_t^2 = \omega + \alpha\big(|z_{t-1}| - E|z_{t-1}|\big) + \gamma z_{t-1} + \beta \ln \sigma_{t-1}^2$
  - z_{t-1}: yesterday's standardised shock; $E|z_{t-1}|$: its mean absolute value ($\sqrt{2/\pi}$ for the Normal distribution); α: size effect; γ: sign effect
  - the logarithm keeps $\sigma_t^2 > 0$ without sign restrictions; leverage effect: $\gamma < 0$; persistence: β
- Both have one parameter more than GARCH(1,1); test $\gamma = 0$ with its t statistic or with an LR test, $\chi^2(1)$

News impact curves: S&P 500 and Bitcoin



- Model curves: σ_t^2 against ε_{t-1}, with σ_{t-1}^2 at the sample variance; dashed: a kernel (Nadaraya–Watson) estimate of $E[r_t^2 \mid r_{t-1}]$, without a model
- S&P 500, GJR: after a fall of 2%, tomorrow's variance is 1.62 times the variance after a rise of 2%; Bitcoin: ratio 1.00, a symmetric curve

TSA_ch5_asymmetry

Asymmetry in four markets

	GJR α	GJR γ	$t(\gamma)$	LR	p	ΔBIC	EGARCH γ	$t(\gamma)$
S&P 500	0.000	0.205	10.0	235.7	< 0.001	-226.8	-0.165	-15.2
BET	0.172	0.043	2.1	4.9	0.026	3.9	-0.028	-2.6
EUR/RON	0.103	0.039	3.0	9.3	0.002	-0.7	-0.040	-2.5
Bitcoin	0.113	-0.013	-0.7	0.7	0.416	7.7	0.015	1.2

- Student-t innovations; robust t statistics; LR $= 2(\ell_{GJR} - \ell_{GARCH}) \sim \chi^2(1)$; ΔBIC: GJR minus GARCH (negative: GJR preferred)
- S&P 500: GJR $\hat{\alpha} = 0.000$: only bad news raises volatility; a very strong leverage effect
- BET: weak, significant by LR, but BIC prefers GARCH; Bitcoin: none; EUR/RON: $r_t > 0$ means a weaker leu, so the "bad news" of GJR are days when the leu strengthens
- Interpretation: leverage is a property of mature equity markets, not of every financial series

TSA_ch5_asymmetry

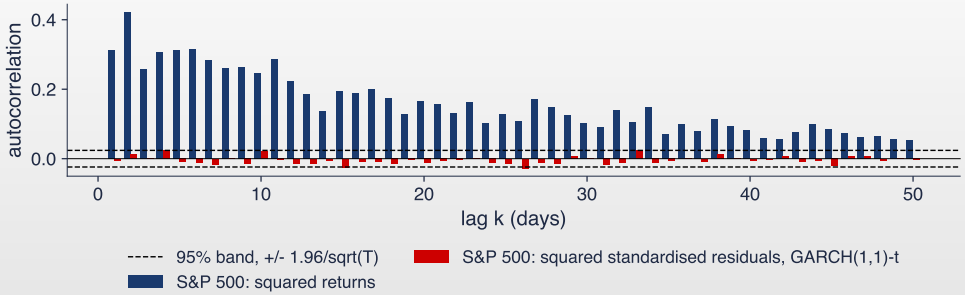
Recap: Asymmetry

- GJR: an extra slope γ for negative shocks; EGARCH: a sign term γz_{t-1} in $\ln \sigma_t^2$
- The news impact curve shows the asymmetry at a glance
- Strong in the S&P 500, weak in the BET, absent in Bitcoin

Checking a fitted model

- If the model is right, the **standardised residuals** $\hat{z}_t = (r_t - \hat{\mu}_t)/\hat{\sigma}_t$ are close to i.i.d. with mean 0 and variance 1
  - Ljung–Box $Q(10)$ on $\hat{z}_t$: no autocorrelation left in the mean (the check of Chapter 2)
  - Ljung–Box $Q(10)$ on $\hat{z}_t^2$ and ARCH-LM on $\hat{z}_t$: no ARCH effects left
  - QQ plot of $\hat{z}_t$: is the assumed distribution right?
- **Sign-bias test** Engle and Ng (1993): regress $\hat{z}_t^2$ on I_{t-1}, $I_{t-1}\hat{\varepsilon}_{t-1}$ and $(1 - I_{t-1})\hat{\varepsilon}_{t-1}$
  - joint test $TR^2 \sim \chi^2(3)$: a rejection means that the sign of past shocks still predicts the variance
- A passed diagnostic does not prove the model right; a failed one shows where it is wrong

What GARCH removes



- S&P 500: ACF of r_t^2: 0.31 at lag 1, still 0.053 at lag 50; ACF of $\hat{z}_t^2$ (GARCH(1,1)-t): -0.005 at lag 1, at most 0.028 in absolute value; band ± 0.024
- Interpretation: four parameters absorb almost all the volatility clustering of 27 years of daily data

TSA_ch5_diagnostics

Diagnostics on real data

S&P 500	$Q(10)$ (p)	$Q(10)$ of squares (p)	ARCH-LM(5) (p)
returns r_t	96.3 (<0.001)	6133 (<0.001)	1684 (<0.001)
$\hat{z}_t$, GARCH-N	21.2 (0.020)	16.4 (0.089)	7.0 (0.222)
$\hat{z}_t$, GARCH-t	21.9 (0.016)	13.6 (0.192)	4.8 (0.438)

- GARCH removes the ARCH effects ($Q(10)$ of squares: from 6133 to 13.6); a little autocorrelation remains in the mean (constant mean; an AR(1) helps, Section 6)
- Sign bias for the GARCH-t of the S&P 500: $\chi^2(3) = 40.7$ (p < 0.001): the asymmetry of Section 7 is missing
- $Q(10)$ of $\hat{z}_t^2$, GARCH-t: BET 8.3 (p = 0.596), Bitcoin 4.7 (p = 0.910), EUR/RON 0.004: suspiciously small (next slide)

TSA_ch5_diagnostics

Case study: the EUR/RON on 6 May 2025

- In April 2025 the reference rate moved by 0.0038% a day on average: the BNR held the leu almost fixed
  - the IGARCH volatility fell to $\hat{\sigma}_t = 0.0078\%$; then the rate rose by 1.20% (two days after the first round of the presidential election) and by 1.21% the next day
- Standardised residual: $\hat{z}_t = 154$, a "154-sigma" day; the next largest is 13.2
  - $Q(10)$ of $\hat{z}_t^2$: 0.004 with this day, 19.2 ($p = 0.038$) without it: one value hides the remaining ARCH effects
- Lesson: a managed rate is not a free market; GARCH cannot foresee a policy decision; always plot $\hat{z}_t$ before trusting a test



The National Bank of Romania, Bucharest

Photo: Britchi Mirela (2018); CC BY-SA 4.0; Wikimedia Commons

TSA_ch5_diagnostics

Nine models for the S&P 500

- AIC Akaike (1974) and BIC Schwarz (1978) as defined on ARCH(q) and its limits; compare models fitted on the same data

	Normal		Student-t		skewed t	
	log-lik.	ΔBIC	log-lik.	ΔBIC	log-lik.	ΔBIC
GARCH(1,1)	−9218.0	648.9	−9062.5	346.7	−9038.4	307.4
GJR-GARCH(1,1)	−9086.5	394.7	−8944.6	119.9	−8905.1	49.6
EGARCH(1,1)	−9075.6	372.9	−8922.1	74.8	−8880.3	0.0

- ΔBIC: the difference from the best model, EGARCH-skewed t
- From the Normal GARCH: asymmetry alone (Normal EGARCH) lowers BIC by 276, heavy tails alone (GARCH-skewed t) by 341, both by 649
- Interpretation: with 6 714 observations even small gains are "significant"; the out-of-sample comparison of Section 9 is the real test

TSA_ch5_diagnostics

Recap: Diagnostics and model choice

- Standardised residuals: no autocorrelation in $\hat{z}_t$ and $\hat{z}_t^2$, the right distribution, no sign bias
- One extreme $\hat{z}_t$ can hide everything: plot before testing
- AIC/BIC on the same data; in-sample fit is not forecasting ability

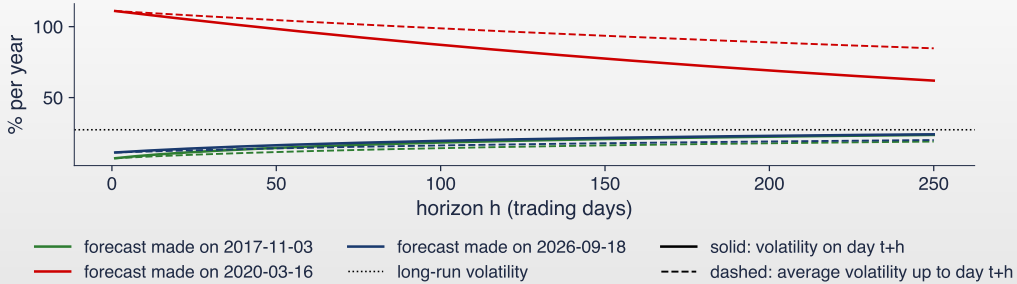
Multi-step forecasts

- One step: $\sigma^2_{t+1} = \omega + \alpha\varepsilon^2_t + \beta\sigma^2_t$ is known at the end of day t
  - two steps: $E_t[\sigma^2_{t+2}] = \omega + \alpha E_t[\varepsilon^2_{t+1}] + \beta\sigma^2_{t+1} = \omega + (\alpha+\beta)\sigma^2_{t+1}$, since $E_t[\varepsilon^2_{t+1}] = \sigma^2_{t+1}$
- By induction: $E_t[\sigma^2_{t+h}] = \bar{\sigma}^2 + (\alpha+\beta)^{h-1}(\sigma^2_{t+1} - \bar{\sigma}^2)$
  - the forecast converges to the long-run variance, like an AR(1) forecast converges to its mean (Chapter 2)
- Variance of the H-day return: the sum of the daily forecasts (the returns are uncorrelated)
  - $\text{Var}_t(r_{t+1} + \cdots + r_{t+H}) = \sum_{h=1}^{H} E_t[\sigma^2_{t+h}]$, not $H\sigma^2_{t+1}$
  - the square-root-of-time rule $\sigma_{t+1}\sqrt{H}$ is too high in a crisis and too low in a calm period

Worked example: the S&P 500 on 18 September 2026

- GARCH(1,1)-t: $\alpha + \beta = 0.9947$, $\bar{\sigma}^2 = 2.972$; forecast for the next day $\sigma^2_{t+1} = 0.509$ (below the long-run level)
- Step by step
  - day 10: $2.972 + 0.9947^9(0.509 - 2.972) = 2.972 + 0.953 \times (0.509 - 2.972) = 0.624$
  - 10-day variance: the sum of the ten daily forecasts $= 5.67$, so the 10-day volatility is 2.38%
  - square-root-of-time rule: $\sqrt{10 \times 0.509} = 2.26\%$: too low, because volatility is expected to rise towards its long-run level

The term structure of volatility



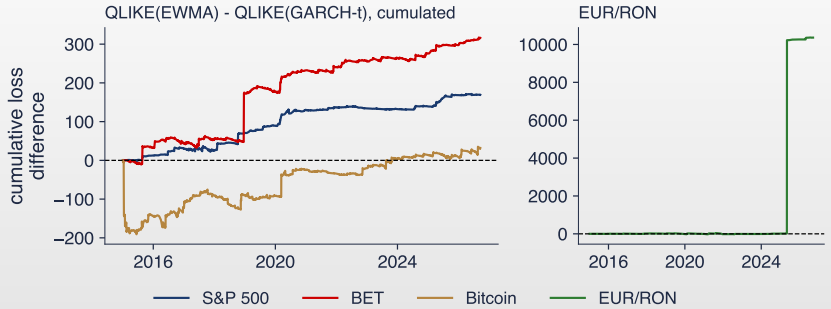
- S&P 500, GARCH(1,1)-t on all data; forecasts made on a calm day (3 November 2017), at the COVID-19 peak (16 March 2020) and on 18 September 2026
- Calm day: 7.3% tomorrow, 19.0% on average over a year; COVID-19 peak: 111.0% tomorrow, still 84.7% on average over a year
- Interpretation: all curves move towards the long-run level (27.3%) at the same slow speed; the slope tells whether we are in a calm period or in a storm

TSA_ch5_forecasts

How to evaluate a volatility forecast

- The true σ_t^2 is never observed: we compare the forecast h_t with a noisy **proxy**, here r_t^2
  - $E_{t-1}[r_t^2] = \sigma_t^2$ (with $\mu \approx 0$), but r_t^2 is very noisy: a low R^2 even for a perfect forecast Andersen and Bollerslev (1998)
- **QLIKE** loss Patton (2011): $L(r_t^2, h_t) = \dfrac{r_t^2}{h_t} + \ln h_t$ (up to a constant); the smaller the mean, the better
  - it ranks forecasts correctly even with a noisy proxy; it is minus the Normal log-likelihood; it punishes forecasts that are too low
- **Out of sample** (Chapter 0): from 2015, each forecast uses only data up to the day before; the models are re-estimated every 250 days
- **DM test** Diebold and Mariano (1995): H_0: $E[d_t] = 0$, with $d_t = L_t^A - L_t^B$ the loss difference of two models A and B
  - $\mathrm{DM} = \bar{d} / \widehat{\mathrm{SE}}(\bar{d})$, approximately $N(0,1)$ under H_0; $\bar{d}$: the mean of d_t; $\mathrm{DM} < -1.96$: A is better
  - $\widehat{\mathrm{SE}}$: HAC (heteroskedasticity and autocorrelation consistent) standard error

GARCH against EWMA, out of sample



- One-day-ahead variance forecasts from 2015 to 18 September 2026: GARCH(1,1)-t (re-estimated every 250 days) against EWMA with $\lambda = 0.94$; a rising curve means that GARCH-t has the smaller QLIKE

TSA_ch5_forecasts

Interpreting the forecast comparison

2015–2026	days	QLIKE GARCH-t	QLIKE GJR-t	QLIKE EWMA	DM t, GARCH-t against EWMA (p)
S&P 500	2 944	0.761	0.733	0.819	-3.69 (<0.001)
BET	2 926	0.680	0.674	0.788	-2.22 (0.027)
EUR/RON	2 939	2.456	3.276	5.982	-1.02 (0.310)
Bitcoin	4 279	3.383	3.401	3.390	-0.20 (0.845)

- S&P 500 and BET: GARCH-t beats EWMA ($t < -1.96$); GJR-t has the smallest QLIKE (the leverage effect pays off out of sample)
- Bitcoin: no difference ($t = -0.20$): its GARCH is an IGARCH, which is almost an EWMA
- EUR/RON: a much smaller mean QLIKE for GARCH, yet $t = -1.02$: 98% of the total difference comes from one day, 6 May 2025, when EWMA forecast a near-zero variance
- Interpretation: QLIKE punishes forecasts that are too low; the DM test protects us from conclusions built on a single day

TSA_ch5_forecasts

Case study: Hansen and Lunde (2005)

- **Question**: does anything beat a GARCH(1,1)? Hansen and Lunde (2005)
  - 330 ARCH-type models (asymmetric, long memory, different innovations), compared out of sample on the Deutsche Mark/US dollar exchange rate and on the returns of one US stock
  - the proxy: realised variance from intraday returns, more precise than r_t^2
- **Finding**: for the exchange rate, no model beat GARCH(1,1) significantly; for the stock, models with a leverage effect did better
  - the same pattern as our table: asymmetry matters for equities, not for currencies or Bitcoin
- **Legacy**: GARCH(1,1) is the benchmark every new volatility model, including machine learning (Chapter 9), must beat out of sample

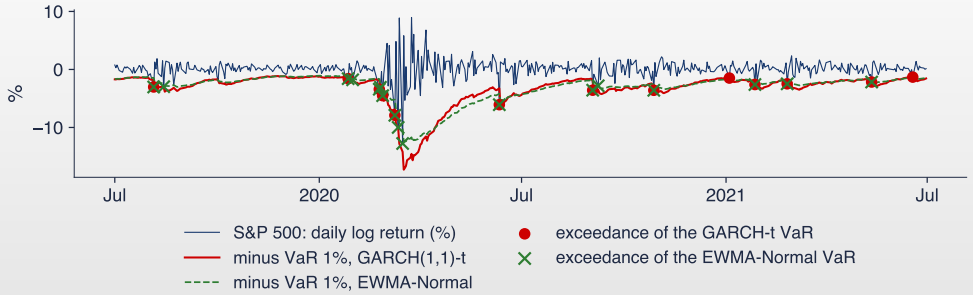
Recap: Variance forecasts

- $E_t[\sigma^2_{t+h}] = \bar{\sigma}^2 + (\alpha+\beta)^{h-1}(\sigma^2_{t+1} - \bar{\sigma}^2)$; H-day variance = sum of the daily forecasts
- Compare forecasts out of sample with QLIKE and the Diebold–Mariano test
- GARCH beats EWMA where volatility mean-reverts

Conditional VaR

- **VaR** (value at risk) at level α: the loss exceeded with probability α, $\text{VaR}_\alpha = -q_\alpha(r_{t+1})$, with q_α the α-quantile
  - here $\alpha = 1\%$: VaR 1%, a loss exceeded on one day in a hundred
- With $r_{t+1} = \mu + \sigma_{t+1} z_{t+1}$: $\text{VaR}_{t+1} = -(\mu + \sigma_{t+1}\, q_\alpha(z))$
  - $q_\alpha(z)$: the α-quantile of the innovation z; Normal: $q_{0.01} = -2.326$
  - standardised t: $q_{0.01} = t_\nu^{-1}(0.01)\sqrt{(\nu-2)/\nu}$, t_ν^{-1}: the quantile function of the Student distribution with ν degrees of freedom
  - the VaR moves every day with σ_{t+1}: high in storms, low in calm periods
- **Worked example**: S&P 500 on 18 September 2026, GARCH(1,1)-t: $\hat{\mu} = 0.079$, $\sigma_{t+1} = 0.714$, $\hat{\nu} = 6.26$
  - $t_{\hat{\nu}}^{-1}(0.01) = -3.100$, $\sqrt{(\hat{\nu}-2)/\hat{\nu}} = 0.825$, so $q_{0.01} = -3.100 \times 0.825 = -2.557$
  - VaR 1% $= -(0.079 + 0.714 \times (-2.557)) = 1.75\%$
  - with Normal z_t: 1.58%; over 10 days (sum of the variances, μ ignored): $2.557 \times \sqrt{5.67} = 6.09\%$

VaR 1% through the COVID-19 crash



- S&P 500, July 2019 – June 2021 (505 days): one-day VaR 1% from GARCH(1,1)-t (re-estimated every 250 days) and from EWMA with Normal z_t
- Exceedances (return below −VaR): GARCH-t 13, EWMA-Normal 17, about 5 expected; both models react, but only after the first large falls of March 2020

TSA_ch5_forecasts

Exceedances 2015–2026

VaR 1%	days	expected	GARCH-t	EWMA-Normal
S&P 500	2 944	29	49 (1.66%)	65 (2.21%)
BET	2 926	29	34 (1.16%)	67 (2.29%)
EUR/RON	2 939	29	35 (1.19%)	63 (2.14%)
Bitcoin	4 279	43	64 (1.50%)	93 (2.17%)

- A correct VaR 1% is exceeded on 1% of the days, and the exceedances do not cluster
- EWMA-Normal: 2.1–2.3% everywhere: the Normal quantile is too small for heavy tails
- GARCH-t: 1.2–1.7%: much closer to 1%; the highest for the S&P 500, where the leverage effect is missing
- Are the differences significant? The test of Kupiec (1995) compares the number of exceedances with a binomial distribution

TSA_ch5_forecasts

Recap: VaR 1% from a GARCH model

- $\text{VaR}_{t+1} = -(\mu + \sigma_{t+1} q_\alpha(z))$: a dynamic volatility and the right quantile
- Normal quantiles underestimate the risk; Student-t is much closer to 1%
- Several series at once (covariances of returns): multivariate GARCH, Chapter 14 (self-study)

Possible contribution of AI

- **Code**: a first draft of a script that runs ARCH-LM, fits GARCH, GJR and EGARCH with `arch` on many series and tabulates the results
- **Explanation**: a second explanation of the ARMA(1,1) form of GARCH, or of an `arch` output table
- **Exploration**: volatility of all BVB stocks, or of the leu against several currencies
- Example prompt
  - `Write Python code that loads daily BET closes, computes log returns in percent, runs the ARCH-LM test with 5 lags, fits AR(1)-GARCH(1,1) with Student-t innovations using the arch package, and plots the annualised conditional volatility.`

Checks you must run

- Units: returns in %, not decimals; rescale a series with a very small variance (EUR/RON) and convert ω back
- The persistence: $\alpha + \beta$ for GARCH, $\alpha + \beta + \gamma/2$ for GJR, β for EGARCH
- Boundary estimates ($\alpha + \beta = 1$): no half-life and no long-run volatility; do not report infinite numbers
- Calendars: Bitcoin trades 7 days a week, the BET five; annualise each series with its own frequency
- Out-of-sample forecasts use only past data; the VaR level is written as VaR 1%
- Every cited reference: it must exist; check the DOI

Key takeaways

- Returns: an almost unpredictable mean and a predictable variance, $r_t = \mu_t + \sigma_t z_t$; test with the ACF of r_t^2 and ARCH-LM
- GARCH(1,1) is an ARMA(1,1) for ε_t^2; $\alpha + \beta$ close to 1; IGARCH and EWMA when $\alpha + \beta = 1$
- Estimate by maximum likelihood with `arch`, report robust SE; Student-t innovations for the tails
- ARMA-GARCH: the mean of Chapter 2 and the variance of this chapter estimated together
- Equity indices need asymmetry (GJR, EGARCH); check $\hat{z}_t$ and $\hat{z}_t^2$ after every fit
- Forecasts revert to the long-run level; compare them out of sample with QLIKE and DM; VaR 1% $= -(\mu + \sigma_{t+1} q_{0.01}(z))$

Key formulas

Quantity	Formula
ARCH-LM	$\hat{\varepsilon}_t^2 = b_0 + \sum_{i=1}^{q} b_i \hat{\varepsilon}_{t-i}^2 + u_t, \quad \text{LM} = nR^2 \sim \chi^2(q)$
ARCH(q)	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2, \quad \bar{\sigma}^2 = \omega/(1 - \sum \alpha_i)$
GARCH(1,1)	$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \quad \bar{\sigma}^2 = \omega/(1-\alpha-\beta), \quad h_{1/2} = \ln 0.5 / \ln(\alpha+\beta)$
EWMA	$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda) r_{t-1}^2$
GJR, EGARCH	$\sigma_t^2 = \omega + (\alpha + \gamma I_{t-1})\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2; \quad \ln\sigma_t^2 = \omega + \alpha(\lvert z_{t-1}\rvert - E\lvert z_{t-1}\rvert) + \gamma z_{t-1} + \beta \ln\sigma_{t-1}^2$
Log-likelihood	$\ell = -\frac{1}{2}\sum_t [\ln 2\pi + \ln\sigma_t^2 + (r_t - \mu_t)^2/\sigma_t^2]$
Forecast	$E_t[\sigma_{t+h}^2] = \bar{\sigma}^2 + (\alpha+\beta)^{h-1}(\sigma_{t+1}^2 - \bar{\sigma}^2)$
QLIKE, VaR 1%	$L = r_t^2/h_t + \ln h_t; \quad \text{VaR}_{t+1} = -(\mu + \sigma_{t+1}\, q_{0.01}(z))$

Self-assessment

- **Question**: $\omega = 0.05$, $\alpha = 0.05$, $\beta = 0.90$. What are the long-run variance and the half-life?
  - **Answer**: $0.05/0.05 = 1$; $\ln 0.5 / \ln 0.95 = 13.5$ days
- **Question**: an ARCH-LM regression with $q = 5$ on 1000 residuals gives $R^2 = 0.004$. Are there ARCH effects at 5%?
  - **Answer**: LM $= 1000 \times 0.004 = 4 < 11.07$: H_0 is not rejected, no evidence of ARCH effects at 5% (which does not prove they are absent)
- **Question**: GJR gives $\hat{\alpha} = 0$ and $\hat{\gamma} = 0.2$. What happens to the variance after a rise of 3%?
  - **Answer**: nothing beyond the decay $\beta\sigma_t^2$: only negative shocks raise the variance
- Next: Chapter 6, VAR models and Granger causality (several series at once)

References (1/3)

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. doi:10.1109/TAC.1974.1100705
- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4), 885–905. doi:10.2307/2527343
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. doi:10.1016/0304-4076(86)90063-1
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69(3), 542–547. doi:10.2307/1925546
- Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2), 143–172. doi:10.1080/07474939208800229
- Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4), 407–432. doi:10.1016/0304-405X(82)90018-6
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. doi:10.1080/07350015.1995.10524599
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007. doi:10.2307/1912773
- Engle, R. F. (2004). Risk and volatility: Econometric models and financial practice. *American Economic Review*, 94(3), 405–420. doi:10.1257/0002828041464597

References (2/3)

- Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1), 1–50. doi:10.1080/07474938608800095
- Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5), 1749–1778. doi:10.1111/j.1540-6261.1993.tb05127.x
- Franke, J., Härdle, W. K., & Hafner, C. M. (2019). *Statistics of Financial Markets: An Introduction* (5th ed.). Springer. doi:10.1007/978-3-030-13751-9
- 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. *Journal of Finance*, 48(5), 1779–1801. doi:10.1111/j.1540-6261.1993.tb05128.x
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. doi:10.2307/j.ctv14jx6sm
- Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3), 705–730. doi:10.2307/2527081
- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873–889. doi:10.1002/jae.800
- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer. doi:10.1007/978-3-031-13584-2
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. otexts.com/fpp3

References (3/3)

- J.P. Morgan/Reuters (1996). *RiskMetrics – Technical Document* (4th ed.). Morgan Guaranty Trust Company. msci.com
- Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84. doi:10.3905/jod.1995.407942
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. doi:10.1093/biomet/65.2.297
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4), 394–419. doi:10.1086/294632
- McLeod, A. I., & Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4(4), 269–273. doi:10.1111/j.1467-9892.1983.tb00373.x
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347–370. doi:10.2307/2938260
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256. doi:10.1016/j.jeconom.2010.03.034
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464. doi:10.1214/aos/1176344136
- Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). Wiley. doi:10.1002/9780470644560