








Modelling Financial Markets

Chapter 5: Models of Conditional Volatility



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

Can We Forecast Conditional Volatility?



Robert F. Engle. **Photo: Jon Demske (2022); CC BY-SA 4.0; Wikimedia Commons**

"These are mean zero, serially uncorrelated processes with nonconstant variances conditional on the past, but constant unconditional variances. For such processes, the recent past gives information about the one-period forecast variance."

— *Engle (1982)*

- S&P 500 (Standard & Poor's): the lag-1 autocorrelation is -0.099 for daily returns and 0.313 for squared returns
  - the sign of a return is nearly unpredictable, its size is not
- ARCH tests (autoregressive conditional heteroskedasticity) reject a constant variance in all six markets of the chapter
  - S&P 500: $Q(10)$ on r_t^2 is 6 135, against a 5% critical value of 18.3
- GARCH(1,1), the generalised ARCH model, with Student-t innovations: persistence $\hat{\alpha} + \hat{\beta} = 0.9948$ and a half-life of 132 days for the S&P 500
  - for Bitcoin and EUR (euro)/RON the estimate hits the bound $\alpha + \beta = 1$: in the model, shocks never fade
- Out of sample, 2016–2026 (2 693 one-day S&P 500 forecasts): GARCH(1,1)-t has a lower QLIKE (quasi-likelihood) loss than EWMA (exponentially weighted moving average) and than the mean of r^2 over the last 20 days
  - Diebold–Mariano $t = 3.47$ and 4.10

▸ Central chart: Out-of-Sample Design and Forecast Paths, 2016–2026

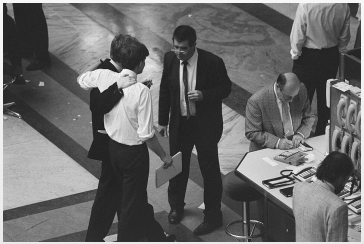
The Idea of the Chapter

- ARCH means AutoRegressive Conditional Heteroskedasticity: variance depends on past shocks
- Core topics
  - Clustering and the ARCH tests; the GARCH recursion, persistence and half-life; estimation with Student-t innovations; forecasts; asymmetry (GJR, EGARCH); residual diagnostics; forecast evaluation with QLIKE, Mincer–Zarnowitz and Diebold–Mariano
- Inference beyond the basic model
  - QMLE (Quasi-Maximum Likelihood Estimator) asymptotics, the delta method near $\alpha + \beta = 1$, tests on the boundary, corrected residual tests, Giacomini–White and the Model Confidence Set
- Two research sections: a quasi score-driven volatility model and an AI-assisted (Artificial Intelligence) discovery loop
- Detailed examples and advanced topics are in Further Topics, after the Conclusions
  - the footer button leads there; the Back button returns to the starting slide

What you should be able to do

- derive the long-run variance, the kurtosis and the half-life of GARCH(1,1)
- compute multi-day variance forecasts and likelihood-ratio (LR) tests on the boundary
- interpret GARCH, GJR and EGARCH estimates in six markets
- interpret out-of-sample comparisons with QLIKE, the Diebold–Mariano (DM) test and the Model Confidence Set
- Further Topics and Appendix: reading for familiarity

Volatility Comes in Bursts

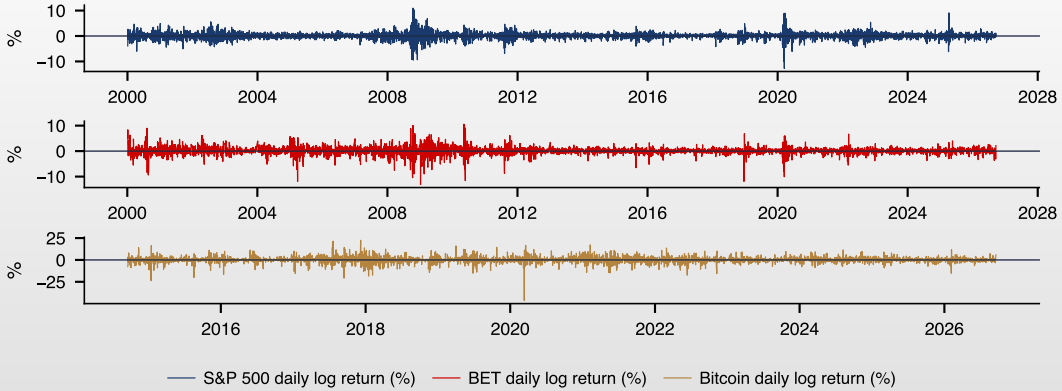


Amsterdam Stock Exchange, 21 October 1987, after the crash of 19 October

Photo: Bart Molendijk / Anefo (1987); CC0; Wikimedia Commons

- Crashes are not confined to isolated days: large moves are followed by large moves
  - "Large changes tend to be followed by large changes, of either sign" Mandelbrot (1963)
- This **volatility clustering** is stylised fact 6 of Chapter 1
- Volatility enters every risk number
  - Value at Risk (Chapter 7)
  - Portfolio weights (Chapter 4)
  - Option prices
- Beyond the forecast: can we conduct valid inference on it?

Volatility Clusters in Every Market



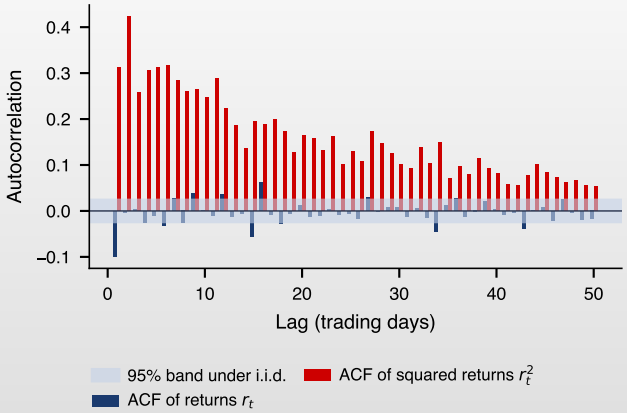
MFM_ch5_arch_effects ▶ Interactive

Calendar and Volatility Episodes

- Daily log returns $r_t = 100\,(\ln P_t - \ln P_{t-1})$ in %, closes up to 18 September 2026
  - P_t: closing price (index level) on day t
- Six series, chosen for different market regimes
  - The S&P 500 (the global benchmark) and the BET (the local market), from 2000
  - BET-TR (the BET with reinvested dividends), from 2014
  - Bitcoin (BTC), from 17 September 2014, 7 days a week
  - EUR/RON (the BNR reference rate, a managed float), from 2005
  - Gold (XAU/USD), a safe-haven asset, from 2000
- Equity indices and gold: weekdays only
  - A heuristic rule: a row that repeats the previous close and shows no trading (reported zero volume, open = high = low = close, or a copy of the previous row) is treated as a provider fill
  - Such rows are dropped, with the reason kept, and checked against the exchange calendar; a genuine unchanged close (a zero return) is kept
  - A missing volume is not treated as zero volume
  - Days dropped: 0 for the S&P 500, the BET and BET-TR, 7 for gold
- In the chart (S&P 500, BET, Bitcoin), quiet years alternate with turbulent months
  - 2008, March 2020, April 2025

MFM_ch5_arch_effects

Predictability of the Sign and of the Size of Returns



- Sample autocorrelation at lag k: $\hat{\rho}_k = \widehat{\text{Corr}}(x_t, x_{t-k})$, computed for $x_t = r_t$ and $x_t = r_t^2$
  - Band $\pm 1.96/\sqrt{n}$ (n: number of returns): 95% range of $\hat{\rho}_k$ if the series were i.i.d.
- S&P 500, lag 1: -0.099 for r_t, 0.313 for r_t^2
- For r_t, 15 of 50 autocorrelations lie outside ± 0.024, but all are small
  - For r_t^2, all 50 do
- Heavy tails too: excess kurtosis (kurtosis minus 3) 10.7 (S&P 500), 11.1 (BET), 12.0 (Bitcoin)
  - The kurtosis requires a finite fourth moment, which Chapter 1 questions
- Object to model: $\sigma_t^2 = \text{Var}(r_t \mid \mathcal{F}_{t-1})$, with $\mathcal{F}_{t-1}$ the information at the close of day $t-1$

MFM_ch5_arch_effects ▶ Interactive

Testing for ARCH Effects (1/2): the Two Tests

- H_0: no clustering, $\mathrm{Var}(r_t \mid \mathcal{F}_{t-1})$ constant
  - Both tests look for autocorrelation in squared returns
- **McLeod–Li** McLeod & Li (1983): the Ljung–Box statistic applied to r_t^2
  - $Q(m) = n(n+2)\sum_{k=1}^{m} \dfrac{\hat{\rho}_k^2}{n-k}$
  - $\hat{\rho}_k$: autocorrelation of r_t^2 at lag k; n: sample size; $m = 10$ lags (two trading weeks)
  - Measures the joint size of the first m autocorrelations; χ^2_{10} under H_0; reject if $Q(10) > 18.3$ (5%)
- **Engle: Lagrange Multiplier (LM) test** Engle (1982)
  - Regress $\hat{\varepsilon}_t^2$ on a constant and $\hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-5}^2$ (5 lags: one trading week)
  - $\hat{\varepsilon}_t = r_t - \bar{r}$: demeaned return, $\bar{r}$: sample mean
  - $LM = nR^2$, with R^2 the coefficient of determination of this regression; H_0: all five slopes are 0
  - χ^2_5 under H_0; reject if $LM > 11.1$ (5%); a large LM means past squared shocks predict today's

Testing for ARCH Effects (2/2): Results

- Every statistic is far above its critical value (all $p < 0.001$)
  - Clustering in equities, Bitcoin, EUR/RON and gold
  - The χ^2 p-values require at least a finite fourth moment of r_t, which the Hill estimates of Chapter 1 put in doubt
- The weakest effect, by $\hat{\rho}_1(r^2)$, is for gold (0.110) and Bitcoin (0.137)
- Q and LM grow with n, so they are not comparable across series of different lengths
  - One-sided score tests: Appendix
- Type of evidence: description (it shows what happened in the data, without testing a cause)

Series	$Q(10)$ on r^2	LM(5)	$\hat{\rho}_1(r^2)$
S&P 500	6 135	1 685	0.313
BET	2 440	982	0.330
BET-TR	714	371	0.171
Bitcoin	213	115	0.137
EUR/RON	2 420	871	0.342
Gold	745	331	0.110

MFM_ch5_arch_effects

The ARCH Model of Engle (1982)

- Setting: $r_t = \mu + \varepsilon_t$, $\varepsilon_t = \sigma_t z_t$, innovations z_t i.i.d. with mean 0 and variance 1
  - μ: mean return; ε_t: shock (unexpected part of the return); σ_t: conditional volatility
  - Conditional variance $\sigma_t^2 = \mathrm{Var}(r_t \mid \mathcal{F}_{t-1})$; unconditional variance $\sigma^2 = \mathbb{E}\sigma_t^2$
- **ARCH**(q): today's variance is a constant plus a weighted sum of the last q squared shocks Engle (1982)
  - $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$
  - q: number of lags; ω: variance floor; α_i: weight of the shock i days ago
  - $\omega > 0$, $\alpha_i \geq 0$: the variance is always positive
- A large shock yesterday (ε_{t-1}^2 large) raises today's variance: clustering by construction
- ARCH(1): unconditional variance $\sigma^2 = \omega/(1-\alpha_1)$, finite if $\alpha_1 < 1$
- Engle received the 2003 Nobel Prize for ARCH (history: Appendix)

From ARCH to GARCH (Bollerslev, 1986)

- GARCH: Generalised AutoRegressive Conditional Heteroskedasticity
  - The variance also depends on its own past
- Problem with ARCH: clustering in real data lasts for months, so q must be large (many parameters)
- **GARCH**(1,1) Bollerslev (1986): $\sigma_t^2 = \omega + \alpha\,\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2$
  - $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$; α: reaction to news; β: memory of past variance
- Substituting backwards: $\sigma_t^2 = \frac{\omega}{1-\beta} + \alpha \sum_{j\geq 1} \beta^{j-1} \varepsilon_{t-j}^2$
  - Valid for $\beta < 1$; j: age of the shock in days, weight $\alpha\beta^{j-1}$
  - An ARCH(∞) with geometrically declining weights, using only three parameters
- GARCH(p,q) adds more lags (p lags of σ^2, q lags of ε^2)
  - In practice GARCH(1,1) is the workhorse Hansen & Lunde (2005)

Model Labels and Innovations

- The model name describes the variance dynamics
  - The suffix describes the law of the standardised innovation
- GARCH-N: GARCH with innovations from the Normal distribution
- GARCH-t: GARCH with Student-t innovations, which allow heavier tails
- GJR-t (or GJR-GARCH-t)
  - Asymmetric Glosten–Jagannathan–Runkle variance (Asymmetry section) with Student-t innovations

GARCH(1,1) as an ARMA Model for Squared Shocks

- Define $u_t = \varepsilon_t^2 - \sigma_t^2$, the surprise in the squared shock ($\mathbb{E}[u_t \mid \mathcal{F}_{t-1}] = 0$)
- Then $\varepsilon_t^2 = \omega + (\alpha + \beta)\,\varepsilon_{t-1}^2 + u_t - \beta\, u_{t-1}$
  - Obtained by adding u_t to both sides of the GARCH recursion and writing $\sigma_{t-1}^2 = \varepsilon_{t-1}^2 - u_{t-1}$
  - An AutoRegressive Moving Average model, ARMA(1,1), in ε_t^2 with autoregressive coefficient $\alpha + \beta$ and moving-average coefficient $-\beta$
- Consequence: the autocorrelations of ε_t^2 decay geometrically, $\rho_k = \rho_1(\alpha + \beta)^{k-1}$
  - $\rho_k = \mathrm{Corr}(\varepsilon_t^2, \varepsilon_{t-k}^2)$: autocorrelation of squared shocks at lag k
  - $\rho_1 = \alpha(1 - \beta^2 - \alpha\beta)/(1 - \beta^2 - 2\alpha\beta)$ if $\mathbb{E}\varepsilon_t^4 < \infty$ (Seminar 5, A6)
  - S&P 500, GARCH-N by Gaussian QMLE ($\hat{\alpha} = 0.120$, $\hat{\beta} = 0.860$): $\rho_1 = 0.356$ and $\rho_{10} = 0.298$, against the sample values $\hat{\rho}_1 = 0.313$ and $\hat{\rho}_{10} = 0.248$
- $\alpha + \beta$ is therefore called the **persistence** of volatility

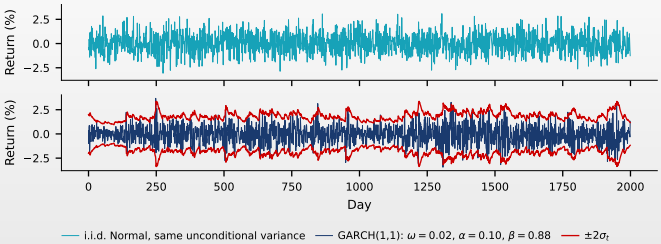
Stationarity and the Unconditional Variance

- **Covariance stationarity**: $\alpha + \beta < 1$; then
  - $\sigma^2 = \mathbb{E}[\sigma_t^2] = \dfrac{\omega}{1-\alpha-\beta}$
- Derivation: take expectations in the recursion, $\mathbb{E}\sigma_t^2 = \omega + (\alpha+\beta)\,\mathbb{E}\sigma_{t-1}^2$, and set $\mathbb{E}\sigma_t^2 = \mathbb{E}\sigma_{t-1}^2$
- **IGARCH** (integrated GARCH): $\alpha + \beta = 1$
  - The unconditional variance is infinite
  - Yet, if $\alpha > 0$ and z_t^2 is non-degenerate, the process is still **strictly** stationary: the condition $\mathbb{E}\ln(\alpha z_t^2 + \beta) < 0$ Nelson (1990) holds by Jensen, since $\mathbb{E}\ln(\alpha z_t^2 + \beta) < \ln(\alpha+\beta) = 0$
- Notation: σ^2 is the model's long-run variance (Seminar 5 writes $\bar{\sigma}^2$)
  - $\hat{\sigma}^2 = \hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$: its estimate
  - s^2: the sample variance of r_t
- Variance targeting: fix $\omega = s^2(1-\alpha-\beta)$, so that the long-run variance equals s^2 and only α, β are estimated

Fourth Moment and Kurtosis

- With Normal z_t, the kurtosis of ε_t is $K = \dfrac{3\left[1-(\alpha+\beta)^2\right]}{1-(\alpha+\beta)^2-2\alpha^2}$
  - $K = \mathbb{E}\varepsilon_t^4/(\mathbb{E}\varepsilon_t^2)^2$: 3 for the Normal distribution, above 3 for heavy tails
  - Finite only if $(\alpha+\beta)^2 + 2\alpha^2 < 1$ (Seminar 5, A4)
  - ARCH(1) is the case $\beta = 0$: $K = 3(1-\alpha^2)/(1-3\alpha^2)$, finite only if $\alpha < 1/\sqrt{3}$: heavy tails even with Normal innovations
- S&P 500, GARCH-N, 2000–2026: $\hat{\alpha} = 0.120487$, $\hat{\beta} = 0.860062$
  - $(\hat{\alpha}+\hat{\beta})^2 + 2\hat{\alpha}^2 = 0.990509$: just below 1, so $K = 12.2$; sample kurtosis of the returns: 13.7 (excess kurtosis 10.7)
- Near the moment boundary, K is very sensitive to rounding the estimates (Appendix)
- Clustering alone generates much of the fat tails
  - Heavy-tailed innovations z_t may contribute further; the gap between 12.2 and 13.7 is not an identified decomposition

Simulation: GARCH Produces Clustering



- Top: i.i.d. Normal returns
- Bottom: GARCH(1,1) with $\omega = 0.02$, $\alpha = 0.10$, $\beta = 0.88$
  - The same unconditional variance (1), and the bands $\pm 2\sigma_t$
- Quiet stretches and bursts appear only in the GARCH path
- Normal innovations, yet fat-tailed returns
  - A mixture of Normal distributions with varying variances has heavy tails
  - Kurtosis on a long path: ▸ Appendix

MFM_ch5_garch_simulation ▶ Interactive

Persistence and Half-Life

- Multi-step forecast: $\mathbb{E}[\sigma^2_{t+h} \mid \mathcal{F}_t] = \sigma^2 + (\alpha+\beta)^{h-1}(\sigma^2_{t+1} - \sigma^2)$
  - h: forecast horizon in days; σ^2_{t+1}: one-day-ahead variance, known at t
  - The gap to the long-run variance shrinks by the factor $\alpha+\beta$ each day
- **Half-life**: number of days after which half of a variance shock has gone
  - $(\alpha+\beta)^h = 0.5 \;\Rightarrow\; h_{1/2} = \dfrac{\ln 0.5}{\ln(\alpha+\beta)}$
- Half-life: 13.5 days for $\alpha+\beta = 0.95$
  - 69 days for 0.99
  - 138 days for 0.995
- Small changes in $\alpha+\beta$ near 1 imply very different half-lives
  - Estimation uncertainty matters (next section; Seminar 5, A2 and B4)

A Toy Check: Multi-Day Risk After a Turbulent Day

- Toy GARCH(1,1), not fitted to any data: $\omega = 0.04$, $\alpha = 0.08$, $\beta = 0.90$
  - Persistence $p = \alpha + \beta = 0.98$; long-run variance $\sigma^2 = 0.04/0.02 = 2$; half-life $\ln 0.5 / \ln 0.98 \approx 34.3$ days
- A turbulent day: $\sigma^2_{t+1} = 4$, twice the long-run level
  - Day 2: $\mathbb{E}[\sigma^2_{t+2} \mid \mathcal{F}_t] = 2 + 0.98 \times (4 - 2) = 3.96$: the gap to σ^2 shrinks by 2% a day
- H-day variance: a geometric sum
  - Daily shocks are uncorrelated
  - So the variance of their sum is the sum of the daily variance forecasts
  - $V_H = \sum_{h=1}^{H} \mathbb{E}[\sigma^2_{t+h} \mid \mathcal{F}_t] = H\sigma^2 + (\sigma^2_{t+1} - \sigma^2)\,\dfrac{1 - p^H}{1 - p}$
  - $H = 5$: $V_5 = 10 + 2 \times (1 - 0.98^5)/0.02 = 10 + 2 \times 4.804 = 19.61$; the square-root-of-time rule, $V_H = H\,\sigma^2_{t+1}$ (H-day volatility is $\sqrt{H}$ times the one-day volatility), gives $5 \times 4 = 20$, too high
- Your turn: redo $H = 5$ after a calm day, $\sigma^2_{t+1} = 1$
  - Answer: $V_5 = 10 - 4.804 = 5.20 > 5$
  - So the rule is now too low, as in the S&P 500 case of the Forecasts section

EWMA and IGARCH: No Mean Reversion

- **Exponentially Weighted Moving Average (EWMA)** J.P. Morgan/Reuters (1996)
  - $\sigma_t^2 = \lambda\,\sigma_{t-1}^2 + (1-\lambda)\,r_{t-1}^2$
  - $\lambda = 0.94$ for daily data and zero-mean returns
  - $\lambda \in (0,1)$: decay factor
  - Weights $(1-\lambda)\lambda^{j-1}$ on past squared returns: the most recent days count most
  - EWMA is an IGARCH(1,1) with $\omega = 0$, $\alpha = 1-\lambda$, $\beta = \lambda$: no parameter to estimate
- Multi-step forecasts at the close of day t differ
  - $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$: expectation given the information at t
  - EWMA: flat, $\mathbb{E}_t\sigma_{t+h}^2 = \sigma_{t+1}^2$ for every h
  - IGARCH with $\omega > 0$: rising, $\mathbb{E}_t\sigma_{t+h}^2 = \sigma_{t+1}^2 + (h-1)\,\omega$
  - Stationary GARCH: pulled towards σ^2 at the rate $\alpha + \beta$
- EWMA serves as a benchmark in the forecast comparison

Maximum Likelihood

- Maximum Likelihood Estimation (MLE) of $\theta = (\mu, \omega, \alpha, \beta)$
  - Given θ, the recursion produces $\sigma_1^2(\theta), \ldots, \sigma_n^2(\theta)$
- Normal log-likelihood: $\ell(\theta) = -\frac{1}{2} \sum_{t=1}^{n} \left[\ln 2\pi + \ln \sigma_t^2 + \varepsilon_t^2 / \sigma_t^2 \right]$
  - $\varepsilon_t = r_t - \mu$
  - Each day contributes the density of ε_t given the past: the likelihood factorises in time
- No closed form: numerical optimisation under $\omega > 0$, $\alpha, \beta \geq 0$, $\alpha + \beta \leq 1$
  - The `arch` package imposes $\alpha + \beta \leq 1$, a closed set: an estimate can sit on the bound (Bitcoin and EUR/RON in the six-market table)
- Start-up: σ_1^2 is a backcast from the first squared residuals, whose effect fades at rate β
- Scale: returns in %, so that the variance is of order 1
  - EUR/RON has a daily standard deviation of about 0.28%, so it is fitted on 10 r_t; on unscaled returns the optimiser stops at a local optimum

Heavy-Tailed Innovations

- **Student-t** Bollerslev (1987): $z_t = \sqrt{(\nu-2)/\nu}\, t_\nu$, unit variance for $\nu > 2$
  - t_ν: a standard Student-t variable, variance $\nu/(\nu-2)$; the factor rescales it to variance 1
  - ν (degrees of freedom): the smaller ν, the heavier the tails; the fourth moment exists only for $\nu > 4$
- **Skewed-t** Hansen (1994): adds an asymmetry parameter $\lambda \in (-1, 1)$
  - $\lambda < 0$: the left tail is longer (large losses more likely than large gains)
- **Generalised Error Distribution (GED)** Nelson (1991): density $\propto \exp(-\frac{1}{2}|z/\lambda_\nu|^\nu)$
  - $\nu > 0$: shape parameter; λ_ν: scale constant
  - $\nu = 2$: the Normal distribution; $\nu = 1$: Laplace; $\nu < 2$: heavier tails (λ_ν gives unit variance)

The Student-t Likelihood

- Day t's contribution to the GARCH-t log-likelihood, the chapter's baseline model ($\varepsilon_t = r_t - \mu$)

$$\ell_t = \ln \Gamma\left(\tfrac{\nu+1}{2}\right) - \ln \Gamma\left(\tfrac{\nu}{2}\right) - \tfrac{1}{2} \ln[\pi(\nu-2)] - \tfrac{1}{2} \ln \sigma_t^2 - \tfrac{\nu+1}{2} \ln\left(1 + \frac{\varepsilon_t^2}{(\nu-2)\sigma_t^2}\right)$$

- Γ: the gamma function
  - $\nu > 2$: degrees of freedom of the unit-variance innovation
- As $\nu \to \infty$, ℓ_t becomes the Normal contribution of the Maximum Likelihood slide
- Why Student-t: a single extra parameter (ν) and symmetric heavy tails

Gaussian QMLE: Definition and Consistency

- **Gaussian QMLE** (Quasi-Maximum Likelihood Estimator) Bollerslev & Wooldridge (1992)
  - Maximise the Normal likelihood even if z_t is not Normal
  - The Normal $\ell(\theta)$ is used as a criterion, not as a claim about the law of z_t
  - θ_0: true parameter value; $\hat{\theta}$: the estimate, the maximiser of ℓ
- **Consistency** (GARCH(1,1)): $\hat{\theta} \to \theta_0$ as $n \to \infty$ Berkes, Horváth & Kokoszka (2003), Francq & Zakoïan (2004)
  - Strict stationarity: $\mathbb{E} \ln(\alpha z_t^2 + \beta) < 0$
  - $\alpha_0 > 0$ (otherwise β is not identified) and $\beta_0 < 1$
  - θ_0 in a compact (closed and bounded) parameter set; z_t^2 non-degenerate (not a constant)
  - No moment condition on ε_t: consistency holds even for IGARCH

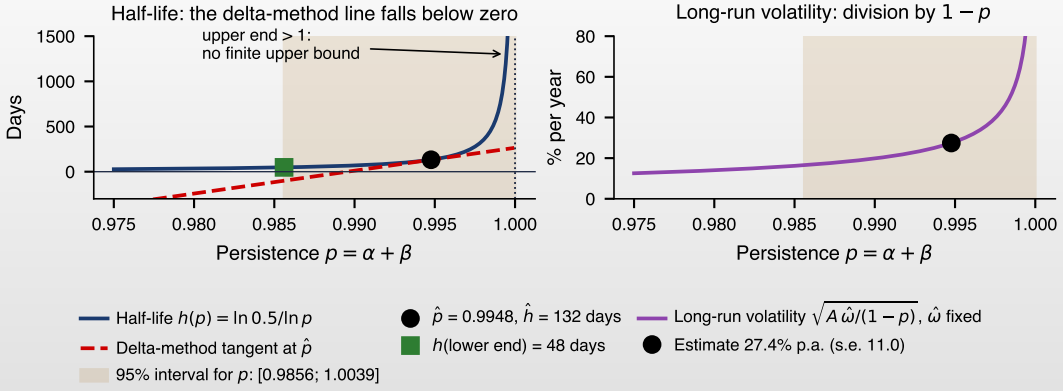
S&P 500: GARCH(1,1)-t Estimates

Parameter	Estimate	Robust SE	Classical SE	Ratio
ω	0.0157			
α	0.122	0.0107	0.0101	1.06
β	0.872	0.0103	0.0096	1.08
ν	6.25	0.48		
$\alpha + \beta$	0.9948			

- Mean $\hat{\mu} = 0.079\%$ per day
  - SE: standard error; Ratio = robust SE / classical SE; full table: Seminar 5, B1
- Large $\hat{\beta}$: the variance memory dominates
  - $\hat{\alpha}$: a moderate news reaction
  - $\hat{\nu} = 6.25$: tails much heavier than Normal
- Half-life 132 days
  - Long-run volatility $\sqrt{251.5\,\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})} = 27.4\%$ p.a. ($A = 251.5$ returns a year)
  - Above the sample 19.2%: dividing by $1-\hat{\alpha}-\hat{\beta}$ makes it imprecise (next slide)

MFM_ch5_garch_estimation

Near-Unit Persistence: What the Delta Method Can and Cannot Do (1/2)



MFM_ch5_seminar ▶ Interactive

Near-Unit Persistence: What the Delta Method Can and Cannot Do (2/2)

- $g(\theta)$: a smooth function of the parameters, with gradient ∇g
  - E.g. the long-run volatility or the half-life
- Delta method: $\widehat{\mathrm{SE}}\, g(\hat{\theta}) \approx \sqrt{\nabla g' \hat{V} \nabla g}$, with $\hat{V}$ the robust covariance of $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$
- For the half-life: $p = \alpha + \beta$ and $h = \ln 0.5 / \ln p$
- Long-run volatility 27.4% p.a., SE 11.0 percentage points
- Half-life: $\partial h / \partial p = 25\,330$ at $\hat{p} = 0.994769$
  - $\widehat{\mathrm{SE}}(\hat{p}) = 0.0047$, so $\widehat{\mathrm{SE}}(\hat{h}) \approx 118$ days
  - $\hat{h} \pm 1.96\, \widehat{\mathrm{SE}}$ includes negative half-lives: the linearisation fails
- Transform the interval for p instead: $[0.9856; 1.0039]$ gives $h \in [48; \infty)$
  - Whether $p = 1$ is plausible needs a boundary test (next slides)

MFM_ch5_seminar

Testing on the Boundary (1/2): a Decision Table

Situation under H_0	**Example**	**Likelihood Ratio (LR):** $LR = 2(\ell_1 - \ell_0)$	**5% critical value**
Interior restriction	GJR (asymmetric GARCH) vs GARCH, $\gamma = 0$ (with $\alpha > 0$)	χ^2_1	3.84
One parameter on the boundary, the others identified and interior	Normal vs Student-t ($1/\nu = 0$); IGARCH ($\alpha + \beta = 1$)	$\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$ Self & Liang (1987), Andrews (2001)	2.71
A nuisance parameter unidentified	No ARCH: $\alpha = 0$ leaves β unidentified	non-standard Andrews (2001)	simulate
Several bounds active at once	GJR with $\alpha_0 = 0$ under H_0, test of $\gamma = 0$	non-standard	simulate

- ℓ_1, ℓ_0: maximised log-likelihoods without and with the restriction
  - $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$: $LR = 0$ with probability $\frac{1}{2}$, otherwise χ^2_1
- GJR: γ is the extra variance reaction to negative shocks (Asymmetry section)
- The LR law depends on the true parameter under H_0, not on the estimate of the alternative: on the S&P 500, GARCH-t fitted under $\gamma = 0$ has $\hat{\alpha} = 0.122 > 0$
  - $\hat{\alpha} = 0$ in GJR affects the Wald t for α, not automatically the LR law for $\gamma = 0$

Testing on the Boundary (2/2): Why IGARCH Is a Boundary Case

- IGARCH is on the boundary only because the fit restricts $\alpha + \beta \leq 1$, as `arch` does
  - Under $H_0 : \alpha + \beta = 1$ the unrestricted estimate would exceed 1 about half of the time; the restriction sets it to 1 and then $LR = 0$: hence the mixture and 2.71
  - Without that restriction, $\alpha + \beta = 1$ lies inside the strictly stationary region, the QMLE is still asymptotically Normal there Lumsdaine (1996), and the LR would be compared with χ^2_1 (3.84)
  - QMLE normality is not enough for a quasi-LR: the χ^2_1 law requires a correctly specified density or a corrected statistic
- A Wald t for an estimate at the bound (e.g. $\hat{\alpha} = 0$) is not $N(0, 1)$
  - A parametric bootstrap from $\hat{\theta}$ (samples simulated from the fitted model) is inconsistent when θ_0 is on the boundary Andrews (2000)
  - Simulate under H_0 instead

Worked Example (1/2): The One-Day Forecast

- S&P 500, GARCH(1,1)-t: $\hat{\omega} = 0.0157$, $\hat{\alpha} = 0.122$, $\hat{\beta} = 0.872$, $\hat{\mu} = 0.079$
  - Persistence $\hat{p} = \hat{\alpha} + \hat{\beta} = 0.994769$
  - With the rounded values shown, $1 - \hat{p} = 0.006$ and the long-run variance would be 2.62 instead of 3.00
- Last day $T = 18.09.2026$: $\hat{\varepsilon}_T = r_T - \hat{\mu} = 0.088$, filtered variance $\hat{\sigma}^2_T = 0.565$
  - Filtered variance: the recursion run through the whole sample with $\hat{\theta}$, read off on the last day T
  - $\hat{\sigma}^2_{T+1} = \hat{\omega} + \hat{\alpha}\hat{\varepsilon}^2_T + \hat{\beta}\hat{\sigma}^2_T = \hat{\omega} + 0.0009 + 0.493 = 0.509$, i.e. 11.3% p.a.
- A quiet last day: $\hat{\sigma}^2_{T+1}$ is far below the long-run level, so the forecasts will rise (next slide)

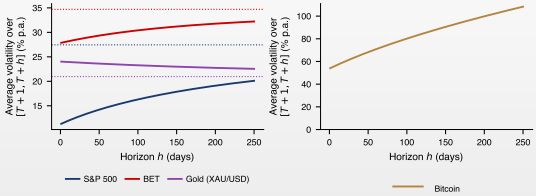
MFM_ch5_garch_estimation

Worked Example (2/2): Ten Days Ahead

- Long-run variance $\hat{\sigma}^2 = \hat{\omega}/(1 - \hat{\alpha} - \hat{\beta}) = 3.00$
  - Day 10: $\hat{\sigma}^2 + (\hat{\alpha} + \hat{\beta})^9(\hat{\sigma}^2_{T+1} - \hat{\sigma}^2)$
  - $= 3.00 + 0.954 \times (0.509 - 3.00) = 0.624$
- 10-day variance $\sum_{h=1}^{10} \hat{\sigma}^2_{T+h} = 5.67$, volatility 2.38%
  - The sum in closed form (toy check in the persistence section): $10\,\hat{\sigma}^2 + (\hat{\sigma}^2_{T+1} - \hat{\sigma}^2)\,(1 - \hat{\rho}^{10})/(1 - \hat{\rho})$
  - $= 29.95 + (0.509 - 2.995) \times 9.768 = 5.67$
  - The square-root-of-time rule gives $\sqrt{10} \times 0.714 = 2.26\%$ ($\hat{\sigma}_{T+1} = 0.714\%$): too low, because $\hat{\sigma}^2_{T+1} < \hat{\sigma}^2$ and volatility is forecast to rise

MFM_ch5_garch_estimation

Volatility Term Structure (1/2)



- Average volatility over the next h days, forecast from 18.09.2026 (GARCH(1,1)-t)
  - Dotted: long-run level
- S&P 500: from 11.3% for the next day to 20.1% on average over 250 days
  - Below its long run today
- Gold falls from 24.0% to 22.5%

MFM_ch5_persistence ▶ Interactive

Volatility Term Structure (2/2)

- Bitcoin and EUR/RON: the estimate hits the bound $\hat{\alpha} + \hat{\beta} = 1$ (IGARCH)
  - With $\hat{\omega} > 0$ (for EUR/RON, $\hat{\omega} \approx 5 \cdot 10^{-8}$ in $\%^2$)
  - So the variance forecast grows linearly by $\hat{\omega}$ per day
- Bitcoin: 54.1% for the next day, 108.3% on average over 250 days (about 8 calendar months)
  - EUR/RON is not in the chart
- This unbounded growth is not a usable forecast
  - In Chapter 8, GARCH forecasts are withheld on days when the fit is not admissible
  - Bitcoin: 1 660 of 3 384 evaluation days
  - EUR/RON: 20 of 4 352

MFM_ch5_persistence

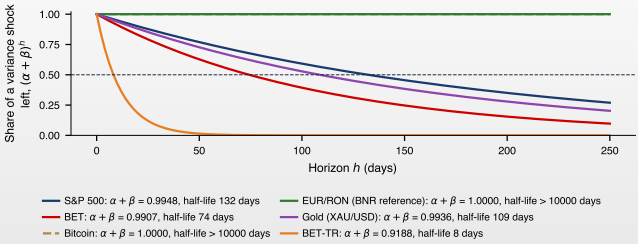
GARCH(1,1)-t in Six Markets

Series	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$	$h_{1/2}$ (days)	$\hat{\nu}$	Long-run vol.
S&P 500	0.0157	0.122	0.872	0.9948	132	6.25	27.4
BET	0.0448	0.192	0.799	0.9907	74	5.01	34.7
BET-TR	0.0755	0.206	0.713	0.9188	8	5.02	15.2
Bitcoin	0.1937	0.109	0.891	1.0000	∞	3.18	–
EUR/RON	0.0000	0.123	0.877	1.0000	∞	4.20	–
Gold	0.0107	0.050	0.944	0.9936	109	4.52	21.0

- Long-run volatility in % p.a., each series annualised with its own frequency
  - –: when $\hat{\alpha}+\hat{\beta} \geq 0.999$
  - EUR/RON: $\hat{\omega}$ in $\%^2$, after the fit on $10\, r_t$
- S&P 500, BET and gold: persistence 0.99–0.995
  - The BET reacts more to news ($\hat{\alpha} = 0.192$) and forgets faster than the S&P 500
- BET-TR covers only 2014–2026: a comparison with the BET needs the same sample
- EUR/RON: on 6 May 2025 a move of $+1.20\%$ after days of $\pm 0.004\%$ gives a standardised residual of 154
- Type of evidence: description

MFM_ch5_garch_estimation

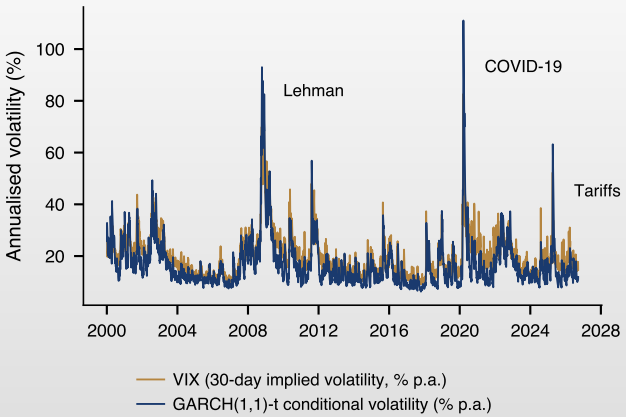
Persistence: How Fast Do Shocks Fade?



- $(\hat{\alpha}+\hat{\beta})^h$: the share of a variance shock left after h days
  - Dashed line: one half
- Half-lives: column $h_{1/2}$ of the six-market table
  - In Bitcoin and EUR/RON shocks never fade in the model
- Level shifts in volatility can mimic high persistence Lamoureux & Lastrapes (1990)
  - Regime switching: Hamilton & Susmel (1994)
  - Report half-lives as intervals (Seminar 5, B4)

MFM_ch5_persistence ▶ Interactive

S&P 500: GARCH Volatility and the VIX (1/2)



- $\hat{\sigma}_t\sqrt{A}$: annualised GARCH(1,1)-t volatility
- VIX: 30-day implied volatility (Volatility Index of the Chicago Board Options Exchange)
- Chart labels
  - Lehman: Lehman Brothers bankruptcy, September 2008
  - COVID-19: Coronavirus Disease 2019, March 2020
  - Tariffs: US (United States) tariff announcement, April 2025

MFM_ch5_sp500_vix ▶ Interactive

S&P 500: GARCH Volatility and the VIX (2/2)

- Correlation of the two series, 2000–2026: 0.89
- $\hat{\sigma}_t$ is filtered with parameters estimated on the full sample (in-sample), not an out-of-sample forecast
- The VIX is higher on average: 19.8% against 17.0%
  - The VIX exceeds GARCH volatility on 84% of the days
- Not a measured variance risk premium
  - The premium compares VIX^2 with a 30-day variance forecast under the real-world probability
  - $\hat{\sigma}_t$ is a one-day volatility
- Type of evidence: association (the variables move together; this does not prove causation)

MFM_ch5_sp500_vix

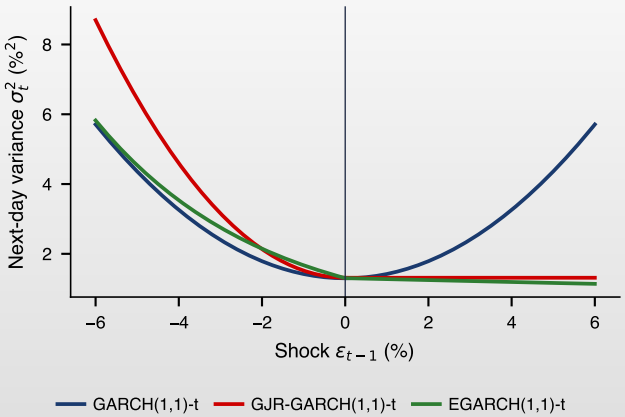
GJR-GARCH (Glosten, Jagannathan and Runkle, 1993)

- **The question**: does volatility rise more after a fall than after a rise?
  - Shocks of the same size (stylised fact 9, Chapter 1)
- GARCH(1,1) cannot see the sign: ε_{t-1}^2 is the same for -3 and $+3$
- GJR: a negative shock gets the extra slope γ Glosten, Jagannathan & 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: slope $\alpha + \gamma$; positivity: $\omega > 0$, $\alpha \geq 0$, $\alpha + \gamma \geq 0$, $\beta \geq 0$ (γ may be negative)
  - Persistence $\alpha + \gamma/2 + \beta$ for symmetric z_t; for skewed z_t replace $1/2$ by $\mathbb{E}[z_t^2 1(z_t < 0)]$, $1(\cdot)$ the indicator (Seminar 5, A3)
- S&P 500: $\hat{\alpha} = 0.000$, $\hat{\gamma} = 0.205$ (t = 10.0), $\hat{\beta} = 0.881$
  - Symmetry test: H_0: $\gamma = 0$ by t or by LR against GARCH(1,1)
- $\hat{\alpha} = 0$ is on the boundary: the t-test of γ is not standard
  - The χ_1^2 law of the LR for $\gamma = 0$ assumes $\alpha_0 > 0$
  - See the decision table for tests on the boundary
- Candidate mechanisms, not tested here
  - Leverage: a falling price raises the debt-to-equity ratio
  - Volatility feedback: higher expected volatility raises required returns, so prices fall now

EGARCH (Nelson, 1991)

- Exponential GARCH (EGARCH) models the log of the variance Nelson (1991)
  - Driven by the size and the sign of yesterday's standardised shock
  - $\ln \sigma_t^2 = \omega + \alpha\left(|z_{t-1}| - \mathbb{E}|z_{t-1}|\right) + \gamma\, z_{t-1} + \beta \ln \sigma_{t-1}^2$
  - $z_{t-1} = \varepsilon_{t-1}/\sigma_{t-1}$: standardised shock; $|z_{t-1}| - \mathbb{E}|z_{t-1}|$: its size above the average size
  - Size effect α, sign effect γ; $\gamma < 0$: negative shocks raise volatility more
- The log form keeps $\sigma_t^2 > 0$ without any sign restriction on the parameters
- Persistence: β; $|\beta| < 1$ gives stationarity of the log-variance $\ln \sigma_t^2$
- Moments of returns need separate conditions: with Student-t innovations the unconditional variance can be infinite even if $\nu > 4$
- S&P 500: $\hat{\alpha} = 0.138$, $\hat{\gamma} = -0.165$ (t $= -15.2$), $\hat{\beta} = 0.979$
- In `arch`, $\mathbb{E}|z|$ is set to $\sqrt{2/\pi}$, its value under the Normal distribution

The News Impact Curve for the S&P 500



- **News impact curve** Engle & Ng (1993): σ_t^2 as a function of ε_{t-1}
  - σ_{t-1}^2 fixed at the sample variance $s^2 = 1.47$ (%^{2})
  - Not at the long-run variance of the model
- Shock of -3%: GJR 3.16, EGARCH 2.75
  - $+3$%: GJR 1.31, EGARCH 1.22
  - GARCH: 2.40 for both
- After bad news the forecast is 2.4 times (GJR) and 2.3 times (EGARCH) that after good news
  - Ratios of levels, not of increases

MFM_ch5_asymmetry ▶ Interactive

Asymmetry in Six Markets

Series	$\hat{\gamma}_{GJR}$	t	LR vs GARCH	p-value	$\hat{\gamma}_{EGARCH}$	t
S&P 500	0.205	10.0	236.2	< 0.001	-0.165	-15.2
BET	0.043	2.1	4.9	0.026	-0.028	-2.6
BET-TR	0.111	2.9	10.3	0.001	-0.063	-3.5
Bitcoin	-0.013	-0.7	0.7	0.416	0.015	1.2
EUR/RON	0.039	3.0	9.3	0.002	-0.040	-2.5
Gold	-0.043	-4.7	26.7	< 0.001	0.038	5.7

- Student-t innovations
  - Robust t
  - $LR \sim \chi^2_1$ (3.84) if the Student-t density is correct
- The robust t is $N(0,1)$ only with the other parameters interior (not for the S&P 500 GJR, where $\hat{\alpha} = 0$)
- Moments: Gaussian QMLE requires $\mathbb{E}z_t^4 < \infty$; a correctly specified Student-t MLE does not; with a wrong density, $\hat{\nu} > 4$ alone does not guarantee the robust errors Fan, Qi & Xiu (2014)
- Bad news weighs more (GJR $\hat{\gamma} > 0$, EGARCH $\hat{\gamma} < 0$)
  - S&P 500 strongly
  - Smaller but significant at 5% for the BET, BET-TR and EUR/RON
  - Bitcoin: none detected
  - Gold: inverted sign
- Type of evidence: association
  - Leverage for equities and a safe-haven role for gold: interpretations consistent with these signs, not tested mechanisms

MFM_ch5_asymmetry

S&P 500: Twelve Models Compared by BIC

ΔBIC	$N(0,1)$	Student-t	Skewed-t	GED
GARCH(1,1)	650.2	347.1	307.9	339.6
GJR-GARCH(1,1)	395.3	119.8	49.6	140.8
EGARCH(1,1)	373.4	74.7	0.0	110.7

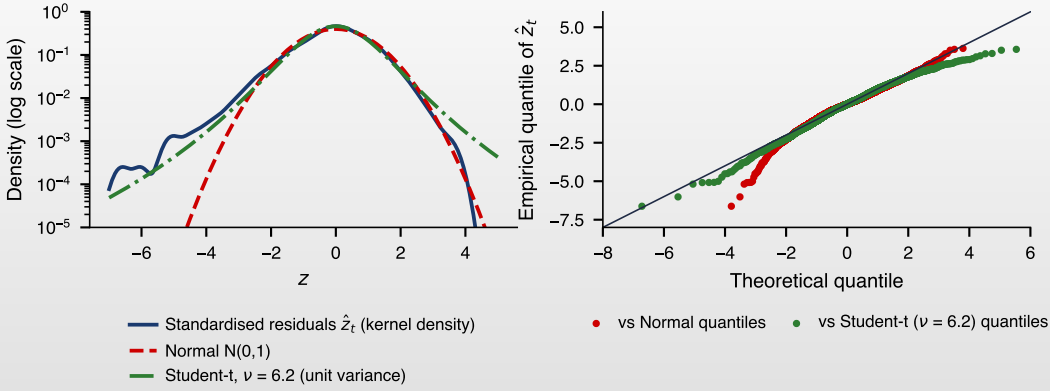
- ΔBIC: Bayesian Information Criterion minus the lowest BIC among the 12 models
  - BIC $= -2\ell + k \ln n$ (k parameters)
  - 0 = preferred model
- Akaike Information Criterion, AIC $= -2\ell + 2k$: we choose by BIC
  - With $n = 6\,717$ ($\ln n \approx 8.8$), BIC penalises each parameter 4.4 times more
  - It avoids over-parameterised models
- Two choices matter
  - Student-t instead of Normal innovations lowers BIC by 276–303
  - Asymmetry lowers it by a further 227–272 (Student-t column, GJR and EGARCH against GARCH)
  - Winner: EGARCH with skewed-t innovations
- BIC ranks in-sample fit, not forecast accuracy
  - The forecast section tests whether these gains survive out-of-sample
- Type of evidence: comparison of procedures (it shows which method works better on the same data, not why)
- Seminar 5, B3 re-estimates the BET row of the asymmetry table and explains the LR, AIC and BIC decisions

MFM_ch5_garch_estimation

Standardised Residuals: What We Check

- **The question**: has the chosen model removed all dynamics from $\hat{z}_t$, and does $\hat{z}_t$ have the assumed density?
- **Standardised residual**: $\hat{z}_t = \hat{\varepsilon}_t / \hat{\sigma}_t$
  - If the model is right, $\hat{z}_t$ is close to i.i.d. with variance 1
- Checks, each with H_0: the model captures that feature
  - Mean equation: Ljung–Box $Q(10)$ on $\hat{z}_t$; variance equation: $Q(10)$ on $\hat{z}_t^2$ and LM(5); distribution: Quantile–Quantile (QQ) plot against the assumed density
- Sign-bias regression Engle & Ng (1993): $\hat{z}_t^2 = b_0 + b_1 S_{t-1}^- + b_2 S_{t-1}^- \hat{\varepsilon}_{t-1} + b_3 S_{t-1}^+ \hat{\varepsilon}_{t-1} + u_t$
  - $S_{t-1}^- = 1$ if $\hat{\varepsilon}_{t-1} < 0$, else 0; $S_{t-1}^+ = 1 - S_{t-1}^-$; u_t: regression error; the size terms use the unstandardised residual $\hat{\varepsilon}_{t-1}$ (eq. 18 in Engle & Ng (1993))
  - b_1 (sign bias): $\hat{z}_t^2$ larger on average after a negative shock, and the sign-bias test is the t of b_1; b_2, b_3: negative and positive size bias; joint test $nR^2 \sim \chi_3^2$ (7.81 at 5%), conservative because it omits the gradient of the fitted variance

S&P 500: Are the Innovations Normal? (1/2)



MFM_ch5_innovations ▶ Interactive

S&P 500: Are the Innovations Normal? (2/2)

- Standardised residuals of GARCH(1,1) with Normal innovations
  - Excess kurtosis falls from 10.7 (returns) to 1.73
  - Skewness -0.53, Jarque–Bera 1 156
- Jarque–Bera $= n[S^2/6 + E^2/24]$, S: skewness, E: excess kurtosis
  - χ^2_2 under Normality, 5% critical value 5.99
  - Normality is rejected by a wide margin
- 16 residuals below -4
  - Expected: 0.21 under the Normal, 8.6 under a Student-t with $\nu = 6.2$
  - The smallest standardised residual: $\hat{z} = -6.6$, on 27 February 2007
- A longer left tail: skewed-t with asymmetry parameter $\hat{\lambda} = -0.155$
  - both from EGARCH with skewed-t innovations: $\hat{\lambda}$ and $\hat{\nu} = 7.5$ degrees of freedom
  - Hence BIC prefers skewed-t innovations

MFM_ch5_innovations

S&P 500: Diagnostics of Five Models

Model	$Q(10)$ on $\hat{z}$	$Q(10)$ on $\hat{z}^2$	LM(5)	Sign bias t	Joint χ^2_3
GARCH-N	21.1	16.2	8.1	3.65	41.6
GARCH-t	21.8	13.5	5.8	3.57	40.6
GJR-t	22.0	10.1	4.4	3.61	28.1
EGARCH-t	27.4	10.5	8.3	2.05	9.9
GJR-skewed-t	19.4	10.6	5.0	3.90	31.9

- 5% critical values: 18.3 (χ^2_{10}), 11.1 (χ^2_5), 7.81 (χ^2_3), $|t| > 1.96$
  - The χ^2 critical values ignore parameter estimation, so the tests are conservative (Further Topics)
- Mean: $Q(10)$ on $\hat{z}$ between 19.4 and 27.4 in every model
  - Above the critical value 18.3: a little autocorrelation that a constant mean ignores
- Variance: $Q(10)$ on $\hat{z}^2$ falls from 6 135 (on r_t^2) to 10–16
  - LM(5) is not significant
  - These naive tests detect no clustering left
- Asymmetry: the joint test rejects all five models at 5%
  - GARCH-t 40.6, GJR-t 28.1, EGARCH-t 9.9 (p-value 0.019, the smallest statistic)
- The sign-bias test (t of b_1) stays above 1.96 in every model (2.05–3.90)
  - With $n = 6\,717$, judge economic size, not only p-values

MFM_ch5_innovations

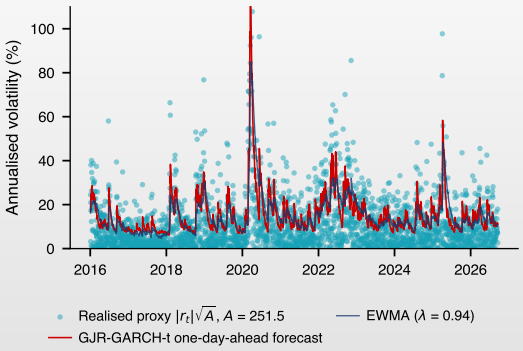
The Proxy Problem and Robust Losses

- **The question**: do the BIC gains of GJR and EGARCH survive out-of-sample?
- Volatility is never observed: we compare forecasts with a noisy **proxy**
  - Here r_t^2, since we use no intraday data
  - Realised variance: Chapter 9
- $r_t^2 = \sigma_t^2 z_t^2$ is unbiased ($\mathbb{E}[r_t^2 \mid \mathcal{F}_{t-1}] = \sigma_t^2$ for $\mu = 0$) but very noisy
  - Even the perfect forecast $h_t = \sigma_t^2$ explains little of the variation of r_t^2 (a small R^2), because z_t^2 is noisy Andersen & Bollerslev (1998)
  - S&P 500: we ignore the mean, since $\hat{\mu}^2 = 0.0062$ against an average daily variance of 1.47 ($\%^2$), a bias below 1%
- Patton (2011): only some loss functions rank forecasts correctly with a noisy proxy
  - Robust: Mean Squared Error (MSE) $(r_t^2 - h_t)^2$ and **Quasi-Likelihood (QLIKE)** $r_t^2/h_t + \ln h_t$, with h_t the variance forecast for day t
  - Not robust: $(|r_t| - \sqrt{h_t})^2$ and $(\ln r_t^2 - \ln h_t)^2$: their ranking depends on the proxy noise
- QLIKE is (up to constants) minus the Normal log-likelihood
  - Its expected value is lowest at $h_t = \sigma_t^2$
  - It penalises under-prediction more

Forecast Evaluation: Mincer–Zarnowitz and Diebold–Mariano

- **Mincer–Zarnowitz (MZ) regression** Mincer & Zarnowitz (1969): regress the proxy on the forecast
  - $r_t^2 = a + b\, h_t + u_t$
- Unbiased forecast: $a = 0$, $b = 1$
  - Joint Wald test with Heteroskedasticity and Autocorrelation Consistent (HAC) errors; χ^2_2 under H_0
  - $b < 1$: forecasts too dispersed (too high when high, too low when low)
- **Diebold–Mariano test** Diebold & Mariano (1995): do two forecasts have the same expected loss?
  - Loss differential $d_t = L(r_t^2, h_t^A) - L(r_t^2, h_t^B)$
  - L: loss function (QLIKE or MSE); h_t^A, h_t^B: forecasts of models A and B; $d_t < 0$: A did better on day t
- Diebold–Mariano (DM) statistic: $DM = \bar{d} / \widehat{\mathrm{SE}}(\bar{d})$
  - $\bar{d}$: mean of d_t over the forecast days; HAC standard error, since d_t is autocorrelated
  - $\approx N(0,1)$ under equal accuracy; $DM < -1.96$: A significantly better; $DM > 1.96$: B significantly better
- The $N(0,1)$ law assumes given forecasts
  - For estimated, nested models (GARCH-t inside GJR-t) it can fail
  - So DM t is read descriptively there (Giacomini–White, GW: Further Topics)

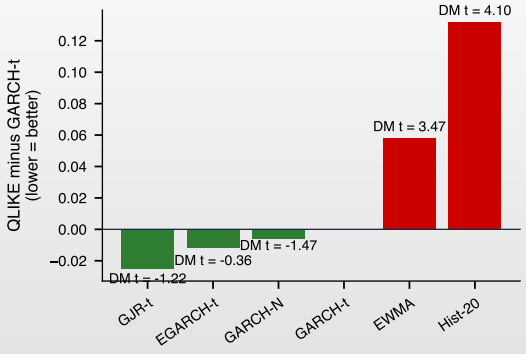
Out-of-Sample Design and Forecast Paths, 2016–2026



- S&P 500: 2 693 one-day forecasts
  - Refit each January on all data since 2000 (expanding window)
  - Daily updates with fixed parameters
- Timing: h_t uses data up to the close of day $t-1$ and is compared with r_t^2
  - Changing r_t must leave h_t unchanged
- Models: GARCH-N, GARCH-t, GJR-t, EGARCH-t
- Benchmarks without estimated parameters: EWMA ($\lambda = 0.94$, the decay factor) and Hist-20, the mean of r^2 over the last 20 days

MFM_ch5_forecast_evaluation ▶ Interactive

S&P 500: Does Anything Beat GARCH(1,1)? (1/2)



- Mean QLIKE minus that of GARCH-t, with the DM t (HAC, 5 lags)
  - Lowest QLIKE: GJR-t (0.731)
- GARCH family: DM t from -1.47 to -0.36, not significant at 5%; descriptive for the nested pairs (GARCH-N and GJR-t against GARCH-t)
  - EGARCH-t is not nested in GARCH-t
  - The DM test requires finite moments and limited dependence of $d_t = r_t^2(1/h_t^A - 1/h_t^B) + \ln(h_t^A/h_t^B)$
  - $\mathbb{E}r_t^4 < \infty$ is not the relevant condition: the forecasts normalise r_t^2; HAC does not repair moments that do not exist
- EWMA and Hist-20 are clearly worse ($t = 3.47$, 4.10)
  - Mean reversion and estimated parameters pay off Hansen & Lunde (2005)
- Type of evidence: comparison of procedures

MFM_ch5_forecast_evaluation ▶ Interactive

S&P 500: Does Anything Beat GARCH(1,1)? (2/2)

- With MSE the ranking changes: EGARCH-t 21.30, GJR-t 21.80, GARCH-N 22.27, GARCH-t 22.76 ($\%^4$)
- MZ: unbiasedness not rejected for GARCH-type forecasts, rejected for Hist-20
  - All statistics: Appendix
- Not rejecting is weak evidence: SE of b is 0.25 for GARCH-t
  - MZ inference requires finite moments of the score $(1, h_t)' u_t$, $u_t = r_t^2 - a - b\, h_t$; even for the ideal forecast, the slope term involves $\mathbb{E} h_t^4$
  - $\mathbb{E} r_t^4 < \infty$ is not enough in general, and HAC does not repair moments that do not exist (heavy tails: Chapter 1)
  - A few crisis days drive b (Seminar 5, B5)

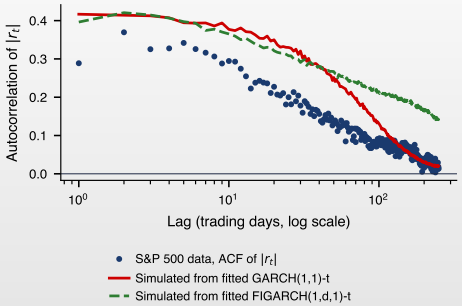
MFM_ch5_forecast_evaluation

GARCH-in-Mean: Is Risk Rewarded?

- **The question**: what does GARCH(1,1) miss, and which extension answers each limit?
- **GARCH-in-Mean (GARCH-M)** Engle, Lilien & Robins (1987): $r_t = c + \kappa\,\sigma_t + \varepsilon_t$ (or $\kappa\,\sigma_t^2$)
  - c: constant; κ: price of risk, the extra expected return per unit of volatility
  - $\kappa > 0$: investors demand a higher expected return when risk is higher (risk-return trade-off)
- S&P 500, GARCH(1,1)-t-M: $\hat{\kappa} = 0.096$ (t = 3.03) with σ_t
  - 0.028 (t = 2.36) with σ_t^2
  - Positive and significant in this sample, but the literature is mixed: Glosten, Jagannathan & Runkle (1993) found a negative relation
- A daily mean effect is small and fragile
  - It is sensitive to the sample and to the asymmetry in variance
- Type of evidence: association

MFM_ch5_long_memory

Long Memory: the Shape of the ACF of $|r_t|$



- S&P 500, ACF (AutoCorrelation Function) of $|r_t|$ (log scale): in the data, 0.289 at lag 1, 0.083 at lag 100 and 0.013 at lag 250
- A simulated GARCH(1,1)-t overstates the ACF at every lag
  - 0.417 at lag 1, 0.131 at lag 100, 0.020 at lag 250
- The problem is the shape: a single $\alpha + \beta$ cannot give both a low level at lag 1 and a slow decay
- **Fractionally Integrated GARCH (FIGARCH)** Baillie, Bollerslev & Mikkelsen (1996)
  - The filter $(1 - L)^d$, with $Lx_t = x_{t-1}$ and $d \in (0, 1)$, gives a hyperbolic decay
- $\hat{d} = 0.60$ (SE 0.06), BIC lower by 34.9
- FIGARCH overstates the memory even more: 0.214 at lag 100, 0.144 at lag 250
- No single-scale model fits all horizons
  - Realised volatility and Heterogeneous AutoRegressive (HAR) models in Chapter 9

MFM_ch5_long_memory ▶ Interactive

Limits of GARCH and the Extensions They Motivate

Limit	Evidence in this chapter	Extension
Level shifts inflate $\alpha + \beta$ Lamoureux & Lastrapes (1990)	BET-TR (2014–2026) far less persistent than the BET (2000–2026): a hint, not proof, as sample and index differ	GARCH-MIDAS (Mixed Data Sampling), a slow long-run component Engle, Ghysels & Sohn (2013); regime switching Hamilton & Susmel (1994)
Reaction lag of one day	COVID-19: the VIX moved the same day, GARCH the next	Realized GARCH, with a realised-variance measurement equation Hansen, Huang & Shek (2012) (Chapter 9)
One outlier drives the update	EUR/RON residual of 154 (May 2025)	GAS (Generalised Autoregressive Score) models, bounded score Creal, Koopman & Lucas (2013) (case study below)
Shape of the ACF of $\lvert r_t \rvert$	GARCH-t and FIGARCH overstate the ACF (previous slide)	HAR on realised variance (Chapter 9)
One asset at a time	portfolios need covariances	Multivariate GARCH and DCC, Dynamic Conditional Correlation (Chapter 6)
Managed prices	EUR/RON: long calm, then jumps	Model interventions as jumps or regimes

- Each extension is judged like GARCH: out-of-sample, robust loss, Model Confidence Set
  - Model equations: Appendix

MFM_ch5_garch_estimation

Three Research Extensions and What They Assume

- **Quasi score-driven volatility**: a replication of a 2023 journal article
  - Builds on: the Student-t density, LR tests and their boundary versions, the score-driven (GAS) update of the extensions section
  - Continued in Seminar 5, C3
- **Volatility cycles (MF2-GARCH)**: a slow component learnt from the forecast errors of GARCH, a replication of a 2025 journal article
  - MF2-GARCH: Multiplicative Factor Multi-Frequency GARCH
  - Builds on: GJR-GARCH, QMLE, the BIC, QLIKE and the Diebold–Mariano test
- **AI for scientific discovery**: an assistant-driven research loop with checks
  - Builds on: the out-of-sample design, QLIKE, Diebold–Mariano and the Model Confidence Set
  - Its checklist supports the AI audit of Seminar 5, C1 and the team project

Score-Driven Volatility: Should Shocks Be Downweighted?

- **Question** Blasques, Francq & Laurent (2023)
  - Must a score-driven volatility model update σ_t^2 with the score of the same density that describes the returns?
- Score-driven (GAS) update Creal, Koopman & Lucas (2013)
  - Tomorrow's parameter moves in the direction that would have made today's observation more likely
  - $f_{t+1} = \omega + \alpha\, S(f_t)\, \partial \log p(y_t \mid f_t)/\partial f_t + \beta f_t$
  - y_t: the return; f_t: the time-varying parameter (here the variance); $p(y_t \mid f_t)$: the density of y_t given f_t
  - $\partial \log p/\partial f_t$: the score; $S(f_t) > 0$: a scaling factor; ω, α, β: as in GARCH
- With a Student-t density, the same parameter ν sets both
  - The tail thickness of z_t
  - How strongly a large shock is downweighted in the update

QSD: Separating Tail Thickness from Shock Downweighting

- **Quasi Score-Driven (QSD) GARCH(1,1)-t**, eq. (30) of the paper
  - $y_t = \sqrt{f_t}\,\epsilon_t$, ϵ_t standardised Student-t with $1/\xi$ degrees of freedom
  - f_t is σ_t^2, ϵ_t is z_t
  - $f_{t+1} = \omega + \varpi X_t + \alpha \dfrac{1+\zeta}{1-2\zeta+\zeta\epsilon_t^2}\,\epsilon_t^2 f_t + \beta f_t$
  - X_t: an exogenous variable, the squared VIX at a daily horizon; ϖ: its coefficient
  - ξ: the tail thickness of ϵ_t ($\xi = 0$: Normal); ζ: the downweighting of large ϵ_t^2 in the update
  - $\zeta = 0$: GARCH(1,1)-t; $\zeta = \xi$: Beta-t-GARCH (score-driven)
  - Why: with $\zeta = \xi = 1/\nu$, multiplying by ν gives $\dfrac{1+\zeta}{1-2\zeta+\zeta\epsilon_t^2} = \dfrac{\nu+1}{\nu-2+\epsilon_t^2}$, the weight in the Student-t score
- Data: 400 S&P 500 stocks, 3 January 1995 (or later) – 28 February 2019
  - Maximum likelihood with an AR(1) mean
  - LR tests of $\zeta = 0$ and $\xi = \zeta$ against χ_1^2 (full method: ▸ Appendix)
  - Probability Integral Transform (PIT) test: $u_t = F(y_t \mid \mathcal{F}_{t-1})$, the fitted distribution function at the observed return, must be i.i.d. uniform on $(0,1)$ if the density is right

QSD: Key Results (Paper, Tables 4 and 6)

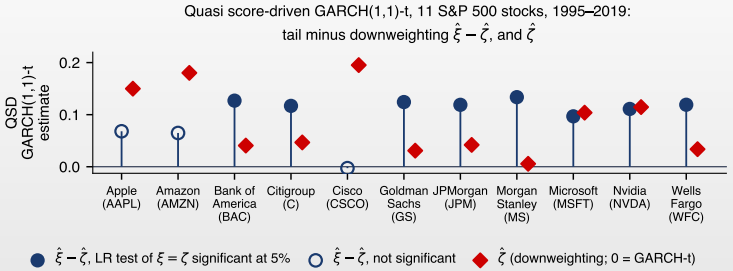
Rejected at 5%, % of the 400 stocks	Student-t (Table 4)	Skewed t (Table 6)
LR, H_0: $\zeta = 0$ (GARCH update)	89.5	98.8
LR, H_0: $\xi = \zeta$ (score-driven; skewed: also $\kappa = \tau$)	53.5	100.0
PIT test: GARCH(1,1)	40.0	33.1
PIT test: Beta-t-GARCH	46.0	72.2
PIT test: QSD GARCH(1,1)	45.7	11.3

- κ: asymmetry of the density (here not the GARCH-M price of risk)
  - τ: asymmetry of the update
  - $\kappa = \tau = 0.5$: symmetric (skewed-t version, Table 6)
- Downweighting large shocks matters (Table 4)
  - $\zeta = 0$ is rejected for most stocks (first row)
  - Every $\hat{\zeta}$ is positive
- Type of evidence: comparison of procedures
  - Beating GARCH-t and Beta-t-GARCH does not make QSD adequate: the PIT test rejects it for 45.7% of the stocks (Table 4)

QSD: Reading the Results (Paper, Figs. 2–6)

- **Beta-t-GARCH downweights too much** (Figs. 2–3)
  - Where $\hat{\xi} - \hat{\zeta}$ is significant, $\hat{\xi} > \hat{\zeta}$ for all stocks but one
  - Median $1/\hat{\zeta}$: 9.48 in QSD, against 5.07 in Beta-t-GARCH, where $\zeta = \xi$
  - Tail thickness and shock downweighting are two different parameters
- **Skewed t** (Figs. 5–6)
  - $\hat{\tau}$ far below 0.5 (median 0.333), while $\hat{\kappa}$ stays near 0.5
  - Leverage enters through the update equation, not through the density
- **Calibration** (Table 6)
  - With skewed-t updates, the PIT test rejects the QSD model for only 11.3% of the stocks, against 72.2% for Beta-t-GARCH

Case Study: Replication on 11 Stocks (cf. Paper, Fig. 2)



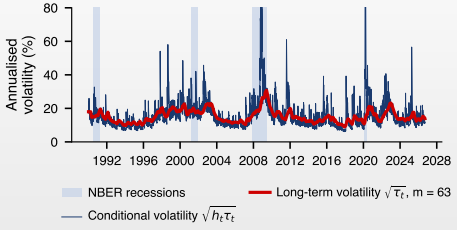
- Circles: $\hat{\xi} - \hat{\zeta}$ (tails minus downweighting) for each stock
  - Filled if the LR test rejects $\xi = \zeta$ at 5%
- Diamonds: $\hat{\zeta}$, the downweighting of large shocks ($\zeta = 0$: GARCH-t)
- $\hat{\xi} > \hat{\zeta}$ for 10 of 11 stocks
  - The exception is Cisco ($\hat{\xi} = 0.193$, $\hat{\zeta} = 0.195$)

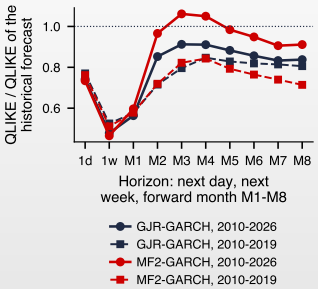
MFM_ch5_quasi_score_driven ▶ Interactive

Volatility Cycles: Can GARCH Learn from Its Own Errors?

- Conrad & Engle (2025): the standardised forecast errors of a one-component GARCH, $r_t^2/\tilde{h}_t$, are predictable
  - the predictor is their average over the last m days: the model underestimates volatility for months in recessions and overestimates it in expansions
- Why it matters: risk forecasts several months ahead depend on the level to which volatility reverts
  - GJR-GARCH always reverts to the same unconditional variance, whatever the phase of the cycle
- MF2-GARCH (Multiplicative Factor Multi-Frequency GARCH): $r_t = \mu + e_t$, $e_t = \sqrt{h_t\,\tau_t}\,Z_t$, with Z_t i.i.d., mean 0, variance 1, symmetric density
  - short-term component: $h_t = (1-\phi) + (\alpha + \gamma\,1\{e_{t-1} < 0\})\,e_{t-1}^2/\tau_{t-1} + \beta h_{t-1}$, with $\phi = \alpha + \gamma/2 + \beta$ and mean 1
  - long-term component: $\tau_t = \lambda_0 + \lambda_1 V_{t-1}^{(m)} + \lambda_2 \tau_{t-1}$ (λ_i: coefficients, not the EWMA factor)
  - $V_{t-1}^{(m)} = \frac{1}{m}\sum_{j=1}^{m} e_{t-j}^2/h_{t-j}$: the average standardised error of h over the last m days
- The paper's numbers, S&P 500, January 1971 to June 2023: the BIC chooses $m = 63$ days (one quarter)
  - persistence: 0.924 in the short term (ϕ) and 0.982 in the long term ($\lambda_1 + \lambda_2$), against 0.982 in GJR-GARCH
  - out of sample, 2010–2019: the RMSE (root mean squared error) for month 1 is 31.5% below that of the historical forecast (ratio 0.685; GJR-GARCH: 0.798)

S&P 500: The Slow Volatility Component and Forecasts Months Ahead





- Left: the conditional volatility $\sqrt{h_t \tau_t}$ and the long-term component $\sqrt{\tau_t}$ of MF2-GARCH with $m = 63$, annualised, 1990–2026
  - $\sqrt{\tau_t}$: maximum 30.8% (4 June 2009), minimum 8.9% (15 December 2017)
- Right: mean QLIKE divided by that of the historical forecast (the mean of r_t^2 over the last 10 years), for the next day, the next week and months 1–8

MFM_ch5_mf2_garch ▶ Interactive

MF2-GARCH: The Paper's Method and the Course Replication

- The paper: Gaussian QMLE, Bollerslev–Wooldridge standard errors, m chosen by the BIC for $m \leq 160$
  - forecasts s days ahead have a closed form (Theorem 2), with $\kappa_z = \mathrm{E}[Z_t^4]$ estimated in sample
  - out of sample, 2010–2023: rolling window, re-estimation every 21 trading days, m re-chosen by the BIC
  - target: realised variance from 5-minute returns plus the overnight return; RMSE and QLIKE losses, relative to the historical forecast
- The course replication: S&P 500, daily log returns, 3 January 1990 to 18 September 2026 (9,245 days)
  - the same specification (MF2-GARCH-rw-m) and the same estimator, on 36 years instead of 52
  - out of sample: 2010–2026, rolling window of 20 years, re-estimation every 21 days, $m = 63$ chosen once, by the BIC on 1990–2009
  - target: the sum of squared daily returns; the course data have no 5-minute returns before 2020
  - QLIKE $= y/f - \ln(y/f) - 1$ (y the target, f the forecast): $y/f + \ln f$ minus $\ln y + 1$, which does not depend on the forecast; the QLIKE ratio omits days with $y = 0$: 1 of 4,203 at the one-day horizon
  - Diebold–Mariano test of MF2-GARCH against GJR on the difference of $y/f + \ln f$, defined also for $y = 0$; Newey–West with as many lags as the target has days
- The 2010–2019 results compare with Table 4 of the paper
  - 2010–2026 includes the March 2020 shock; in the tests of Table 5, the paper gives zero weight to days with realised variance above the 99th percentile

MF2-GARCH: Forecast Gains by Horizon and Period

- In sample, the BIC again chooses $m = 63$; the log-likelihood rises by 57.1 over GJR-GARCH, the BIC falls by 96
  - fast component: $\hat{\phi} = 0.935$ (half-life 10 days); GJR-GARCH: 0.978 (32 days)
  - slow component: $\hat{\lambda}_{\mathbf{1}} + \hat{\lambda}_{\mathbf{2}} = 0.989$, but $\mathbb{E}\tau_t$ follows an AR(63) recursion; its dominant root, 0.9966, corresponds asymptotically to halving in about 204 days
  - a shock to a single daily error raises τ_t for 63 days; the effect then halves after another 137 days
  - $\hat{\lambda}_{\mathbf{1}} = 0.067$ (standard error 0.051), $\hat{\lambda}_{\mathbf{2}} = 0.922$ (0.059); in the paper: 0.112 (0.054) and 0.870 (0.062)
- 2010–2019, relative QLIKE for month 8: 0.714 for MF2-GARCH, 0.806 for GJR-GARCH (DM t -2.66)
  - for month 1: 0.578 and 0.576 (DM t 0.07); relative RMSE for month 1: 0.721 and 0.762
  - in the paper, relative QLIKE for month 8: 0.859 and 0.918
- 2010–2026: the QLIKE ranking reverses from month 1 to month 8 (month 8: 0.911 against 0.838, DM t 0.99)
  - the relative RMSE stays lower for MF2-GARCH for months 1–5 (month 1: 0.916 against 0.960, DM t -1.16)
- Type of evidence: a forecast comparison; the paper's direction holds at long horizons on 2010–2019, not on 2010–2026
- The limits
  - a target built from daily returns is noisier than realised variance, so the DM tests have less power
  - the return variance is finite only if $\Gamma_m < 1$ (Theorem 1 of the paper), a condition on top of $\phi < 1$ and $\lambda_{\mathbf{1}} + \lambda_{\mathbf{2}} < 1$: $\hat{\Gamma}_{\mathbf{63}} = 0.883$ on the full sample, between 0.881 and 0.950 in the 201 rolling windows
  - $\hat{\alpha} = 0.000$, on the boundary (in the paper 0.003): the usual standard errors do not apply to this parameter Andrews (2001)
  - QMLE and losses in variance units require a finite fourth moment, which Chapter 1 questions; $\hat{\kappa}_z = 5.03$

Open Questions and the Discovery Loop (1/2)

- Questions
  - do machine-learning or realised-measure forecasts beat a well-specified GARCH for daily equity variance, out of sample?
  - does the answer survive once the number of models compared is accounted for?
- Literature
  - Christensen, Siggaard & Veliyev (2023): machine-learning methods with little tuning are competitive with HAR-type models
  - Hansen, Lunde & Nason (2011): the Model Confidence Set is the right object when many models compete
- Formal claims, $d_t = L(r_t^2, h_t^A) - L(r_t^2, h_t^B)$, each with its own null and rule
  - Equality of A and B: $H_0 : \mathbb{E}d_t = 0$, rejected at 5% if $|t_{DM}| > 1.96$
  - Superiority of A: $H_0 : \mathbb{E}d_t \geq 0$, rejected at 5% if $t_{DM} < -1.645$
  - Membership of the Model Confidence Set concerns the whole set of models, not the pair A–B
  - Stability: a separate test of equal $\mathbb{E}d_t$ across pre-registered subperiods

Open Questions and the Discovery Loop (2/2)

- The loop
  - An AI assistant drafts the literature map, competing hypotheses, code, a robustness plan and a referee-style critique
  - The researcher checks every Digital Object Identifier (DOI, a permanent publication identifier), reproduces every number and fixes the plan before the first run
- Checks specific to volatility forecasts
  - Look-ahead: h_t must not move when r_t changes
  - A proxy-robust loss (QLIKE, MSE)
  - Model mining: Model Confidence Set or a Bonferroni correction (the 5% level divided by the number of tests)
  - Scale and convergence of the optimiser
  - DOIs or `arch` function arguments invented by the assistant (reusable prompts and the full checklist: Appendix)

Mini-Case and a Team-Project Seed

- A typical assistant claim: "GJR-t beats GARCH-t because of the leverage effect"
  - d_t = QLIKE(GJR-t) − QLIKE(GARCH-t), 2 693 days, 2016–2026: $\bar{d} = -0.025$, DM $t = -1.22$
- An auxiliary hypothesis, not a necessary implication of leverage: the gain sits after down days
  - Regression on the full calendar: $d_t = a + b\,1(r_{t-1} < 0) + u_t$, HAC with 5 lags
  - $\hat{b} = 0.019$ (SE 0.041, $t = 0.46$): no evidence of a difference between days after a fall (1 223) and the others (1 470)
  - By period: 2016–2020 $t = -0.02$, 2021–2026 $t = -4.06$
  - Descriptive only (nested models, expanding window); calibrated follow-up: the same regression in a fixed-window GW test
- Type of evidence: comparison of procedures
- **Project seed**
  - Do machine-learning forecasts enter the Model Confidence Set together with GARCH, GJR and EGARCH for the S&P 500 and the BET, 2016–2026?
  - Replicate the QLIKE table first; pre-register models, losses, splits and error rate
  - Log AI use in `AI_USE.md` and every caught error in `AI_ERRORS.md` (full brief: ▸ Appendix)

Synthesis: a Decision Workflow

- **1. Detect**
  - $Q(10)$ on r_t^2 and the LM test
  - Clustering appears in all six markets; persistence is close to 1, and for Bitcoin and EUR/RON the fit hits the IGARCH bound
- **2. Fit**
  - GARCH, GJR, EGARCH with Normal, Student-t, skewed-t innovations; returns in %
  - Check convergence and parameters at a bound
- **3. Diagnose**
  - Level, squares and signs of $\hat{z}_t$, and its largest values
  - Naive tests are conservative
- **4. Forecast**
  - The recursion gives multi-step variances
  - The square-root-of-time rule ignores mean reversion
- **5. Compare**
  - Out-of-sample: QLIKE, GW with a rolling window and the Model Confidence Set
  - GARCH beats EWMA and Hist-20; the GARCH family is hard to separate
- **6. Report uncertainty**
  - Robust SE (valid if $\mathbb{E}z_t^4 < \infty$) and boundary-aware tests
  - Half-lives as intervals $[h_{lo}; \infty)$ when $\alpha + \beta \approx 1$

Check Yourself, and Next: Chapter 6

Retrieval questions

- Why does the square-root-of-time rule overstate the 10-day risk after a turbulent day?
- Why is the LR test of Normal against Student-t innovations compared with 2.71 and not 3.84?
- Why can a symmetric interval for the half-life contain negative values?
- Why is a gap between the VIX and GARCH volatility not a measured variance risk premium?

Next chapters

- Chapter 6: multivariate volatility and dynamic correlations (DCC)
- Chapters 7–8: Value at Risk and expected shortfall with GARCH volatility, and their backtesting
- Chapter 9: realised volatility and long memory
- Seminar 5: the variance recursion step by step
  - GARCH-t and GJR on real data
  - An AI answer to audit
  - Bootstrap tests and forecast comparisons

Check Yourself: Short Answers

- **Square-root-of-time rule**
  - After a turbulent day $\sigma^2_{t+1} > \sigma^2$, the long-run variance; the daily forecasts decay towards σ^2, so their sum is below $H\sigma^2_{t+1}$
  - Seminar 5, A1
- **2.71 instead of 3.84**
  - Under H_0, $1/\nu = 0$ is on the boundary: in about half of the samples the estimate sits at the bound and $LR = 0$, so the law is $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$, whose 5% critical value is 2.71
- **Negative half-lives**
  - $h = \ln 0.5 / \ln p$ explodes as $p \to 1$; the delta method replaces it by a straight line, so $\hat{h} \pm 1.96\,\widehat{\mathrm{SE}}$ ignores the curvature; map the interval for p instead
  - Seminar 5, A2
- **VIX against GARCH**
  - The right comparison: VIX^2, the 30-day risk-neutral expectation of variance, against a 30-day GARCH variance forecast under the real-world probability; $\hat{\sigma}_t$ is a one-day volatility, so horizon and probability both differ

The Practical Decision

- **What I choose**: a GARCH-family model with Student-t innovations; for equity indices, GJR-t
  - GJR-t has the lowest out-of-sample QLIKE (0.731) and an MCS (Model Confidence Set) p-value of 1.000
  - EWMA and Hist-20 are clearly worse (DM $t = 3.47$ and 4.10)
- **Under which conditions**
  - The GJR-t persistence is below 1: $\hat{\alpha} + \hat{\gamma}/2 + \hat{\beta} = 0.9835$ for the S&P 500, 0.9897 for the BET
  - Active restrictions are reported: for the S&P 500, $\hat{\alpha} = 0$ hits the bound $\alpha \geq 0$, so the standard error of $\hat{\alpha}$ is not the usual one
  - Student-t MLE standard errors assume the correct density; otherwise robust errors require finite moments of the score, which $\hat{\nu} > 4$ alone does not guarantee
  - $Q(10)$ on $\hat{z}_t^2$ and LM(5) detect no clustering left
- **What result would make me drop it**
  - The estimate hits the bound $\hat{\alpha} + \hat{\beta} = 1$, as for Bitcoin and EUR/RON: the variance forecast then grows by $\hat{\omega}$ per day
  - The model leaves the 90% Model Confidence Set, or loses to EWMA in QLIKE out of sample

Gaussian QMLE: Asymptotic Normality and Standard Errors

- **Asymptotic normality** of $\vartheta = (\omega, \alpha, \beta)$ (pure GARCH, known mean): add ϑ_0 interior and $\kappa_z = \mathbb{E}z_t^4 < \infty$
  - $\sqrt{n}(\hat{\vartheta} - \vartheta_0) \to N\left(0, (\kappa_z - 1)J^{-1}\right)$, $J = \mathbb{E}[g_t g_t']$, $g_t = \partial \ln \sigma_t^2 / \partial \vartheta$
  - κ_z: kurtosis of z_t (3 if Normal); g_t: gradient of $\ln \sigma_t^2$; J: its mean outer product; interior: no constraint binds
- Sandwich $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$ estimates $(\kappa_z - 1)J^{-1}$
  - The inverse Hessian $\hat{A}^{-1}$ estimates $2J^{-1}$
  - A: minus the expected Hessian (second derivatives) of ℓ/n; B: variance of the score (first derivatives) per observation
  - Robust / classical standard error (SE) $= \sqrt{(\kappa_z - 1)/2}$: 1 under Normality, above 1 with heavy tails
- Variants (estimated mean, IGARCH, explosive case): Appendix

QMLE in Our Data: When the Theory Holds and When It Fails

- S&P 500, GARCH(1,1) with Normal likelihood: residual kurtosis $\hat{\kappa}_z = 4.73$
  - Theory: robust / classical SE $= \sqrt{(4.73-1)/2} = 1.37$ (Seminar 5, A5); observed: 1.37 for $\hat{\alpha}$, 1.37 for $\hat{\beta}$
- Bitcoin: GARCH-t gives $\hat{\nu} = 3.18 < 4$ degrees of freedom
  - Under that law $\mathbb{E}z_t^4 = \infty$
  - The condition for Gaussian-QMLE normality fails
  - Gaussian QMLE residual kurtosis does not settle: 19.8 on the first half, 14.6 on the full sample
  - Full sample, robust / classical SE: 2.51 for $\hat{\alpha}$, 2.01 for $\hat{\beta}$, against the formula value 2.61, which assumes $\kappa_z < \infty$
- Hall & Yao (2003): then the Gaussian QMLE converges slower than $\sqrt{n}$
  - To a non-Gaussian limit
  - Its robust t-statistics are not valid
- Remedies
  - A heavy-tailed quasi-likelihood with its scale adjustment Fan, Qi & Xiu (2014)
  - Or the Student-t MLE (bounded score: no fourth moment needed if the law is right)
  - With bootstrap inference near bounds

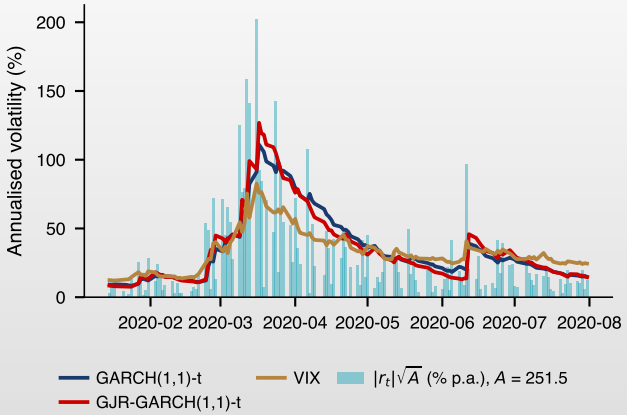
MFM_ch5_inference

Boundary Tests on Our Data

- Normal vs Student-t innovations, GARCH(1,1)
  - $LR = 311.9$ (S&P 500), 1 082.0 (Bitcoin; Seminar 5, B6)
  - Far above 2.71: heavy tails are decisive; the mixture matters only in borderline cases
- IGARCH, $H_0: \alpha + \beta = 1$ (GARCH-t): $LR = 1.09$ (S&P 500), 1.14 (BET)
  - S&P 500, p-value: 0.30 (χ^2_1), 0.15 (mixture), 0.17 (parametric bootstrap under H_0 on $\hat{\alpha} + \hat{\beta}$, 998 converged re-estimates of 999 simulated samples)
  - BET, p-value: 0.28, 0.14 and 0.17 (Seminar 5, B8)
- S&P 500, bootstrap under H_0: 53% of the re-estimates have $\hat{\alpha} + \hat{\beta} \geq 0.999$
  - At the bound $\alpha + \beta = 1$, as the mixture predicts
  - Conclusion: a unit root in variance cannot be rejected for the S&P 500 or the BET

MFM_ch5_seminar

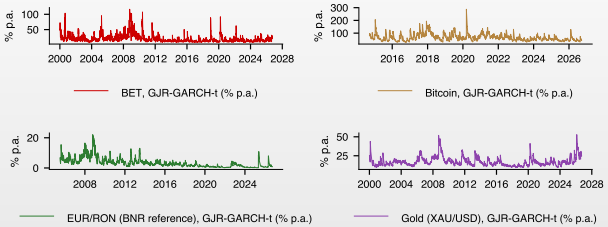
Zoom: The COVID-19 Crash, 2020



- COVID-19, January–July 2020: $|r_t|$ annualised, GARCH-t and GJR-GARCH-t volatility, VIX
- 24 February ($r_t = -3.41\%$): the VIX went from 17.1% to 25.0% the same day
  - GARCH-t from 12.2% to 22.5% only the next day
  - GARCH reacts, markets price at once
- After $r_t = -12.8\%$ on 16 March, peaks on 17 March: GARCH-t 111%, GJR-t 127%
  - The GJR peak is higher because the shocks were negative
  - Full-sample (2000–2026) parameters
  - A forecast with parameters estimated before 2020 (as in Seminar 5, B5) gives a different peak

MFM_ch5_sp500_vix

BET, Bitcoin, EUR/RON and Gold



- BET: 2008 dominates, now 26.4%
  - Bitcoin: 50–100% is normal, the highest peak is March 2020
- EUR/RON: a managed calm with bursts in 2022 and 2025, poorly described by a daily GARCH
  - Gold: a new high of 53% in 2026
- Bitcoin and EUR/RON: the fit is on the IGARCH bound, so the paths are descriptive only (term structure)

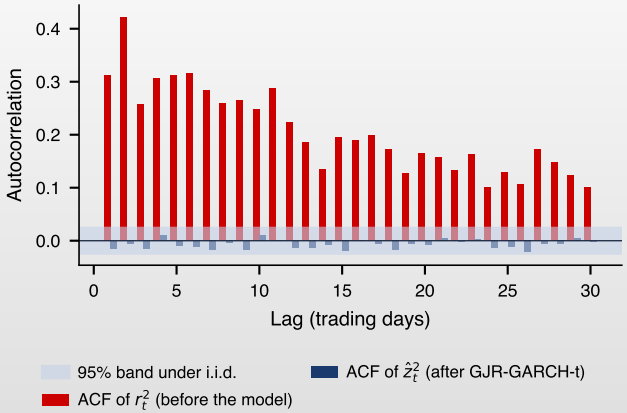
MFM_ch5_markets

Residual Diagnostics Have Estimated Parameters

- $\hat{z}_t$ depends on $\hat{\theta}$: the Ljung–Box statistic on $\hat{z}_t^2$ is not exactly χ^2_m Li & Mak (1994)
  - The fitted variance has already absorbed part of the dependence at small lags
- S&P 500, Gaussian QMLE of GARCH(1,1), $Q(10)$ on $\hat{z}_t^2$
  - Naive Ljung–Box: 16.2, p-value 0.094: not rejected at 5%
  - Corrected for the estimated parameters: 19.0, p-value 0.041: rejected at 5%, the verdict flips
- The naive test is conservative
  - Under H_0 it rejects less often than the nominal level, so it also has less power
  - The joint sign-bias test carries the same caveat (the correction matrix: Appendix)

MFM_ch5_inference

Diagnostics: Squared Returns versus Squared Residuals



- S&P 500, lags 1–30: autocorrelations of r_t^2 (before the model) and of $\hat{z}_t^2$ from GJR-GARCH-t
- After the model they sit inside the i.i.d. band
  - No remaining variance dynamics are detected at these lags
- The band is the naive one: with estimated parameters it is too wide at small lags (previous slide)

MFM_ch5_innovations

Diagnostics in Other Markets: a Decision Table

Market	Residual pattern	Possible failure	Next check
BET	$Q(10)$ on $\hat{z} = 116.6$	autocorrelation in the mean (thin trading, Chapter 2)	AutoRegressive model AR(1) for the mean
Bitcoin	$Q(10)$ on $\hat{z}^2 = 4.7$, joint sign bias 1.3: no rejection	$\hat{\nu} = 3.18 < 4$: $\mathrm{Var}(z_t^2) = \infty$, so the χ^2 calibration is not justified	tests on $\lvert\hat{z}_t\rvert$
EUR/RON	excess kurtosis 3 190	one residual of 154 (May 2025) dominates	inspect the largest residuals; a bounded-score model
Gold	EGARCH: $Q(10)$ on $\hat{z}^2 = 48.6$	clustering left despite a good BIC	choose by diagnostics, not by BIC alone

- Always look at the largest residuals: one observation can dominate every diagnostic

MFM_ch5_innovations

Comparing Forecasting Methods (1/3): the Limit of DM and the Giacomini–White Test

- **The question**: how do we compare forecasts of nested models with estimated parameters?
- **The limit of the DM test** West (1996), Clark & McCracken (2001)
  - With estimated parameters, nested models and an expanding window, the loss differential degenerates under H_0
  - DM is no longer $N(0,1)$
- **The Giacomini–White test** Giacomini & White (2006) tests the forecasting **method**
  - The model plus a rolling window of $W = 1\,000$ days
- The estimation noise does not vanish, so the test also applies to nested models
  - Under the conditions of the theorem: finite moments, limited dependence, a non-singular Ω
- $GW = n\,\bar{Z}^\top \hat{\Omega}^{-1} \bar{Z} \to \chi^2_2$ under $H_0 : \mathbb{E}[d_{t+1} \mid \mathcal{F}_t] = 0$, with $Z_{t+1} = (1, d_t)^\top d_{t+1}$
  - d_{t+1}: loss differential; $\bar{Z}$: mean of Z_{t+1}; $\hat{\Omega} = n^{-1} \sum Z_{t+1} Z_{t+1}^\top$: the uncentred second moment; n: number of forecasts
- Reject at 5% if $GW > 5.99$

Comparing Forecasting Methods (2/3): the Model Confidence Set

- **The question**: which models remain candidates for "the best" when we compare them all?
- With six models, the comparisons with GARCH-t are five tests
  - The **Model Confidence Set (MCS)** Hansen, Lunde & Nason (2011) compares all 15 pairs
- Multiple tests raise the rate of false rejections
- The MCS drops the worst model until equal accuracy is no longer rejected
- The final set contains the best model with probability $\geq 1 - \alpha$, here $\alpha = 10\%$
- Stationary bootstrap: resamples blocks of random (geometric) length
  - Mean 10 days, so that serial dependence is kept

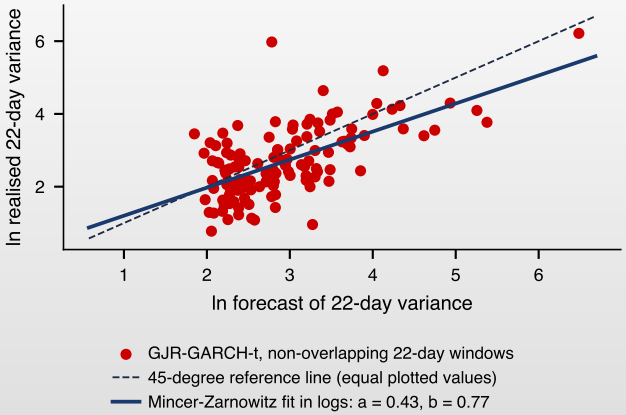
Comparing Forecasting Methods (3/3): Results for the S&P 500

- **Giacomini–White**: rolling window, 2 693 forecasts, each model against GARCH-t
  - Rejects only EWMA and Hist-20
- **Model Confidence Set (MCS) 90%**: expanding window, QLIKE, stationary bootstrap
  - Keeps the four GARCH models, eliminates EWMA and Hist-20
- Same conclusion as the DM test
  - GW, with a rolling window, has calibrated inference under the conditions of the theorem
  - The MCS (expanding window) carries the same caveat as DM
  - Full table and the MCS statistic: Appendix

Model	GW p	MCS p
GARCH-N	0.058	0.588
GARCH-t	–	0.341
GJR-t	0.285	1.000
EGARCH-t	0.202	0.588
EWMA	0.004	0.008
Hist-20	< 0.001	0.002

MFM_ch5_inference

One-Month Forecasts: Mincer–Zarnowitz in Logs (1/2)



- GJR-t forecast of the 22-day variance against the realised sum of 22 squared returns
  - 122 non-overlapping windows
  - Regression $\ln Y = a + b \ln \hat{h} + u$, Y: realised sum, $\hat{h}$: forecast
  - R^2 in logs 0.39
- Slope in logs $b = 0.77$ (SE 0.07): descriptive only
  - Since $\mathbb{E}[\ln Y \mid \mathcal{F}] \neq \ln \mathbb{E}[Y \mid \mathcal{F}]$
  - $b < 1$ does not prove over-dispersed forecasts

MFM_ch5_forecast_evaluation

One-Month Forecasts: Mincer–Zarnowitz in Logs (2/2)

- In levels b is only 0.66: two COVID-19 windows dominate
- Hence we compare the 22-day forecasts with QLIKE Patton (2011)
  - GARCH-N 4.12, GARCH-t 4.13, GJR-t 4.14, EWMA 4.26 (lower = better)
- At one month, the GJR asymmetry no longer pays off

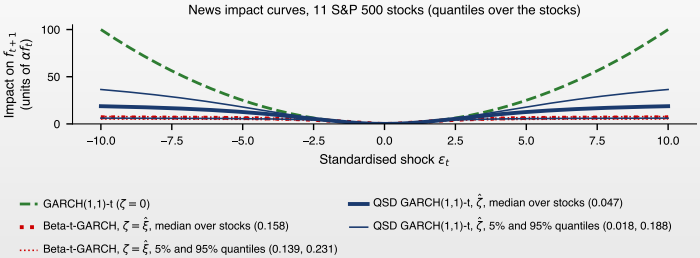
MFM_ch5_forecast_evaluation

QSD: Course Sample and Results

- Course sample: 11 large stocks (the paper: 400)
  - 1995 (or later) – February 2019, 4986–6081 returns each
  - $X_t = \mathrm{VIX}_t^2/252$
- $\zeta = 0$ rejected for 9 of 11 (paper: 89.5%)
  - $\xi = \zeta$ rejected for 8 of 11 (paper: 53.5%)
- Median $1/\hat{\zeta}$: 21.4 (QSD) against 6.32 (Beta-t-GARCH)
  - paper: 9.48 against 5.07
  - 11 selected survivors: imprecise and not representative

MFM_ch5_quasi_score_driven

QSD: News Impact Curves (cf. Paper, Fig. 3)



- Updating functions at quantiles of the 11 estimates, at a common αf_t
  - The contribution of ϵ_t to f_{t+1}, in units of αf_t
  - GARCH-t adds $\epsilon_t^2\ \alpha f_t$
- A shock of 5 standard deviations adds 25 (GARCH-t), 12.6 (median QSD), 6.2 (median Beta-t-GARCH)
  - At 10 standard deviations: 100, 18.8, 7.0 (bound $(1+\zeta)/\zeta = 7.3$)
- The QSD downweighting lies between GARCH-t and Beta-t-GARCH

MFM_ch5_quasi_score_driven

Data Used in This Chapter

Series	Owner / source	Start	Obs.	Obs. per year	Vol. (% p.a.)	Excess kurt.
S&P 500	S&P Dow Jones Indices	2000	6 717	251.5	19.2	10.7
BET	Bucharest Stock Exchange	2000	6 670	249.8	22.1	11.1
BET-TR	Bucharest Stock Exchange	2014	2 992	249.7	15.4	17.8
Bitcoin (BTC/USD)	crypto exchanges (composite)	2014	4 384	365.3	66.6	12.0
EUR/RON reference rate	BNR	2005	5 352	252.4	4.4	15.1
Gold (XAU/USD)	spot market	2000	6 949	260.2	17.6	7.7

- Equity indices and gold: weekdays only
- Vol.: sample volatility annualised with each series' observations per year

MFM_ch5_arch_effects

Back

One-Sided Score Tests and the Davies Problem

- The alternative is one-sided ($\alpha_i \geq 0$): the two-sided χ^2 test wastes power
  - Lee & King (1993): one-sided score test built on the sum of the score components, $N(0,1)$ under H_0, rejects in the upper tail only
- Against GARCH(1,1) the null $\alpha = 0$ leaves β unidentified (Davies problem)
  - The ARCH(q) LM statistic is also the score test against GARCH(p,q): the same statistic serves both alternatives (Lee & King (1993))

Back

Robert Engle and the 2003 Nobel Prize



Robert F. Engle (2022)

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



Clive W.J. Granger (2008)

Photo: Olaf Storbeck (2008); CC BY-SA 2.0; Wikimedia Commons

- Nobel Prize 2003
  - Engle "for methods of analyzing economic time series with time-varying volatility (ARCH)"
  - Shared with Clive Granger (cointegration)
- Nobel lecture: Engle (2004), on risk and volatility in financial practice
- Tim Bollerslev, Engle's doctoral student, generalised ARCH to GARCH in 1986

Back

Kurtosis: Rounding and a Long Simulated Path

- S&P 500, GARCH-N: the moment condition is 0.990509, so $K = 12.2$ has the denominator $1 - 0.990509$
  - Rounding to $\hat{\alpha} = 0.120$, $\hat{\beta} = 0.860$ gives 0.9892 and $K = 11.0$: report such functions from unrounded estimates
- Simulated GARCH(1,1) of the simulation slide, one million days
  - Excess kurtosis 2.47 (theory 3.06); lag-1 autocorrelation of ε_t^2: 0.26
  - The sample kurtosis converges slowly: its precision depends on the eighth moment, infinite for these parameters

Back

QMLE: Variants of the Theorem

- Estimated mean: the score gains the term $\varepsilon_t \, (\partial \mu_t / \partial \theta) / \sigma_t^2$
- The covariance is no longer $(\kappa_z - 1) J^{-1}$: the mean block also depends on the skewness of z_t
- IGARCH ($\alpha + \beta = 1$): strict stationarity still holds Lumsdaine (1996)
  - The QMLE remains consistent and asymptotically Normal
- Explosive case, $\mathbb{E} \ln(\alpha z_t^2 + \beta) > 0$
  - Jensen & Rahbek (2004) estimate (α, β) with ω fixed, since ω cannot be estimated consistently

Back

The Li–Mak Correction for Squared-Residual Autocorrelations

- Squared-residual autocorrelations $\hat{r}_k$: $\sqrt{n}\,\hat{r} \to N(0, C)$, $C = I - H' J^{-1} H/(\kappa_z - 1)$ Li & Mak (1994)
  - $\hat{r} = (\hat{r}_1, \ldots, \hat{r}_m)'$; I: identity matrix; C: the corrected covariance, smaller than I on the diagonal; g_t, J: as in the QMLE theorem
  - $H_k = \mathbb{E}[g_t(z_{t-k}^2 - 1)]$ (k-th column of H): how the fitted variance absorbs lag-k dependence; the ARMA rule χ^2_{m-p-q} does not apply
  - The mean μ is estimated too: C comes from the joint covariance of $\hat{r}$ and the QMLE score for $(\mu, \omega, \alpha, \beta)$, including the cross terms with $\mathbb{E}z_t^3$
- S&P 500, Gaussian QMLE of GARCH(1,1): $C_{1,1} = 0.78$, $C_{5,5} = 0.93$, $C_{10,10} = 0.97$
  - The correction is largest at small lags, where $\hat{\alpha}$ has already fitted the dependence
  - The corrected statistic $n\hat{r}' C^{-1} \hat{r}$ is χ^2_{10} under H_0

MFM_ch5_inference

Back

S&P 500 Forecast Evaluation, All Statistics

Model	QLIKE	MSE	DM t vs GARCH-t	MZ b	MZ R^2	Wald p
GARCH-N	0.750	22.3	−1.47	0.97	0.250	1.00
GARCH-t	0.756	22.8	–	0.88	0.238	0.80
GJR-t	0.731	21.8	−1.22	0.84	0.276	0.68
EGARCH-t	0.745	21.3	−0.36	1.24	0.293	0.71
EWMA	0.814	24.7	3.47	0.89	0.172	0.69
Hist-20	0.888	26.1	4.10	0.68	0.152	0.03

- Mean losses on 2 693 days
  - DM on QLIKE, HAC with 5 lags
  - Negative DM t: the model beats GARCH-t
- MZ: $r_t^2 = a + b\,h_t$; Wald p: test of $a = 0$, $b = 1$

MFM_ch5_forecast_evaluation

Back

Giacomini–White and the MCS: Full Table and Settings

Model	QLIKE (W)	t uncond.	GW	p	MCS p
GARCH-N	0.750	1.86	5.70	0.058	0.588
GARCH-t	0.743	–	–	–	0.341
GJR-t	0.737	−0.19	2.51	0.285	1.000
EGARCH-t	0.745	0.06	3.19	0.202	0.588
EWMA	0.814	3.19	11.31	0.004	0.008
Hist-20	0.888	3.87	16.35	< 0.001	0.002

- GW: the statistic of the main text, χ^2_2 under H_0 (5.99 at 5%)
  - t uncond.: the unconditional test (instrument 1 only)
  - A Diebold–Mariano t on the rolling-window losses
- Rolling window $W = 1\,000$ days, re-estimated each January, 2 693 forecasts
  - Tests against GARCH-t: unconditional t uses HAC with 5 lags
  - GW uses the uncentred second moment, without HAC
- MCS on the expanding-window forecasts: QLIKE
  - Statistic T_R: the largest pairwise $|t|$ of loss differences
  - Stationary bootstrap (mean block 10, 10 000 draws)

MFM_ch5_inference Back

Score-Driven, Realized and Component Models

- **Score-driven (GAS)** Creal, Koopman & Lucas (2013): $\sigma^2_{t+1} = \omega + \alpha\, s_t + \beta\, \sigma^2_t$
  - s_t: the scaled score of the Student-t density with respect to σ^2_t
  - The score is bounded in z_t: an outlier such as the EUR/RON residual of 154 moves the variance far less than in GARCH
- **Realized GARCH** Hansen, Huang & Shek (2012): adds a measurement equation linking σ^2_t to a realised variance
  - Uses intraday information and reacts faster; the bridge to Chapter 9
- **GARCH-MIDAS** (GARCH with Mixed Data Sampling, long-run component) Engle, Ghysels & Sohn (2013): $\sigma^2_t = \tau_t\, g_t$
  - A slow long-run component τ_t driven by macro data (mixed frequencies)
  - g_t: a unit-mean GARCH(1,1) component for the short-run moves
  - A moving long-run level absorbs the level shifts that push $\hat{\alpha} + \hat{\beta}$ towards 1

Back

Case Study, Full Method (1/2)

- **Why it matters**: GARCH is driven by ε_t^2 whatever the density
  - So an isolated extreme shock can produce an excessive volatility response
  - The response lasts days or weeks (Section 1 of the paper)
  - With a left-skewed density, a pure score-driven model raises volatility more after positive returns: the opposite of the leverage effect (Example 2)
  - QSD models, eq. (3): $f_{t+1} = \omega + \alpha\,\psi(y_t, X_t, f_t, \theta) + \beta f_t$, ψ the scaled score of an updating function ρ that need not equal $\log p$; theory: consistency, asymptotic normality, LR and LM tests with χ^2 limits (Section 3)
- **Estimation**: $\zeta = 0$ is kept inside the parameter space by eqs. (31)–(32)
  - $\zeta \in (-1, 1/2)$
  - $\Psi(x) = x\,(1 - e^{-cx})/(1 + e^{-cx})$ applied to the weight, $c = 1000$
  - PIT test Diebold, Gunther & Tay (1998) with the Pierce (1982) correction for parameters estimated by maximum likelihood: reject if $|\sum_t \hat{u}_t - T/2| / \sqrt{T(1/12 - B\,\mathcal{I}(\theta)^{-1}B^\top)} > 1.96$
  - $B = \lim_{T\to\infty} T^{-1}\mathbb{E}[\partial \sum_t u_t / \partial\theta^\top]$, $\mathcal{I}$: the Fisher information per observation, T: the sample length; it detects a shifted PIT mean, not every departure from uniformity

Back

Case Study, Full Method (2/2)

- **Data**: 400 stocks (Section 5)
  - Stocks with fewer than 4000 observations or with convergence problems are dropped
  - Skewed variant, eq. (47): a skewed Student-t with asymmetry κ for ϵ_t and separate (ζ, τ) in the update; $\tau < 0.5$ gives a leverage effect whatever κ (Section 5.2)

Case Study, an Example and the Limits Stated by the Authors

- Centene, 4330 days: $\hat{\xi} \approx 0.26$ (about 4 degrees of freedom) in all three models
  - But $\hat{\zeta} = 0.048$ (SE 0.027)
  - LR 8.81 ($p = 0.003$) for $\xi = \zeta$ and 8.42 ($p = 0.004$) for $\zeta = 0$ (Table 5)
  - After a crash day in March–April 2008, GARCH-t volatility stays well above Beta-t-GARCH
  - QSD lies between (Fig. 4)
- **Limits** stated by the authors
  - Beating GARCH-t and Beta-t-GARCH does not make the Student-t QSD model adequate: the three models fail the PIT test for 40.0–46.0% of the stocks (Table 4)
  - In the unmodified recursion (30), $\zeta < 0$ or $\zeta > 1/2$ does not define a proper volatility model, and $\zeta = 1/2$ with $\varpi = 0$ is degenerate (Section 4.1)
  - The $N(T/2, T/12)$ law of $\sum_t u_t$ holds only for known parameters; with estimated ones the PIT test needs the correction above (Section 5.1)
- **Open extensions** (Seminar 5, C3)
  - Does the bounded QSD update beat GJR-GARCH-t in QLIKE out of sample, especially after March 2020?
  - Does it tame the EUR/RON residual of 154?

Back

The Discovery Loop: Reusable Prompts (1/2)

- Tools are interchangeable
  - An AI assistant (Claude, ChatGPT, Gemini, GitHub Copilot)
  - Literature engines (Semantic Scholar, Elicit, Google Scholar)
- **Literature**
  - *"List out-of-sample comparisons of GARCH, HAR and machine-learning variance forecasts since 2015, with DOI, asset, loss function, test and horizon."*
  - Resolve every DOI on Crossref; check that the loss is robust to a noisy proxy Patton (2011)
- **Hypotheses**
  - *"For the claim that asymmetry improves one-day forecasts, list competing mechanisms (leverage, volatility feedback, a few crisis days) and what each predicts for d_t given the sign of r_{t-1} and by subperiod."*
- **Data and code**
  - *"Write Python with `arch` that re-estimates GARCH-t and GJR-t each January on an expanding window from the course daily price files and stores one-day variance forecasts; no parameter may use data after the forecast origin."*
  - Replicate first: QLIKE 0.756 (GARCH-t) and 0.731 (GJR-t), DM $t = -1.22$

Back

The Discovery Loop: Reusable Prompts (2/2)

- **Robustness**
  - *"Write a pre-registered plan: splits by the sign of r_{t-1} and by subperiod, MSE besides QLIKE, the 22-day horizon, the HAC lag, and a Bonferroni or Model Confidence Set correction."*
  - A random-split placebo is calibrated only if it reassigns blocks of days, preserving volatility clustering
- **Critique**
  - *"Act as a referee: could the proxy r_t^2, one crisis year or the HAC lag create my result? Give the test that would expose each."*
- **Report**: all splits and losses tried, adjusted p-values, code, seeds and package versions
  - A specification chosen after seeing results is exploratory and reported as such
- **Influential periods**: a crisis year can dominate $\bar{d}$
  - Report subperiods and state in advance whether any year may be excluded

Back

The Team-Project Seed in Full

- Realised-variance (HAR) branch: 40 large American stocks
  - The course data set of 5-minute realised variances, January 2015 – January 2026
  - There are no intraday data for the S&P 500 index or the BET, so this branch is run stock by stock
- Replicate first the QLIKE table and the mini-case splits
- Deliverables
  - A one-page pre-registration: models, losses, splits, error rate and what would falsify the claim
  - Code on the course market data, the Model Confidence Set at 10%, a referee-style self-critique; `AI_USE.md` and `AI_ERRORS.md`
- Code of the mini-case: Quantlets/Ch_05/ai_discovery_case.py

Back

Further Resources

- Books
  - Franke, Härdle & Hafner (2019), chapters on AutoRegressive Integrated Moving Average (ARIMA) and GARCH models for financial time series
  - Francq & Zakoïan (2019): the mathematical theory of GARCH models
- Quantinar courses
  - Statistics of Financial Markets
  - Applied Time Series Analysis with Python
- Code
  - All charts and tables: Quantlets/Ch_05; Python package `arch` by Kevin Sheppard (documentation)
- Related work: entropy as a tail-risk signal beyond variance
  - Pele, Lazar & Dufour (2017): entropy of intraday EUR/JPY returns, Value at Risk and expected shortfall
  - Pele & Mazurencu-Marinescu-Pele (2019): intraday Bitcoin entropy and daily Value at Risk
  - Pele, Lazar & Mazurencu-Marinescu-Pele (2019): tail entropy and expected shortfall

Back

Acronyms used in this material (1/2)

- **ACF**: AutoCorrelation Function
- **AI**: Artificial Intelligence
- **AIC**: Akaike Information Criterion
- **AR**: AutoRegressive (model)
- **ARCH**: AutoRegressive Conditional Heteroskedasticity
- **ARIMA**: AutoRegressive Integrated Moving Average (model)
- **ARMA**: AutoRegressive Moving Average (model)
- **BET**: Bucharest Exchange Trading (index)
- **BET-TR**: BET Total Return (index)
- **BIC**: Bayesian Information Criterion
- **BTC**: Bitcoin
- **BNR**: Banca Națională a României — *the National Bank of Romania*
- **COVID-19**: COronaVIrus Disease 2019
- **DCC**: Dynamic Conditional Correlation
- **DOI**: Digital Object Identifier
- **DM**: Diebold–Mariano (test)
- **EGARCH**: Exponential GARCH
- **EUR**: euro, international currency code
- **EWMA**: Exponentially Weighted Moving Average
- **FIGARCH**: Fractionally Integrated GARCH
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GARCH-M**: GARCH-in-Mean
- **GARCH-MIDAS**: GARCH with Mixed Data Sampling (long-run component)
- **GARCH-N**: GARCH with Normal innovations
- **GAS**: Generalised Autoregressive Score (model)

Acronyms used in this material (2/2)

- **GED**: Generalised Error Distribution
- **GJR**: Glosten–Jagannathan–Runkle (asymmetric GARCH)
- **GJR-GARCH**: Glosten–Jagannathan–Runkle GARCH
- **GW**: Giacomini–White (test)
- **HAC**: Heteroskedasticity and Autocorrelation Consistent (standard errors)
- **HAR**: Heterogeneous AutoRegressive (model)
- **IGARCH**: Integrated GARCH
- **LM**: Lagrange Multiplier
- **LR**: Likelihood Ratio (test)
- **MCS**: Model Confidence Set
- **MF2-GARCH**: Multiplicative Factor Multi-Frequency GARCH
- **MLE**: Maximum Likelihood Estimation
- **MSE**: Mean Squared Error
- **MZ**: Mincer–Zarnowitz (regression)
- **NIC**: News Impact Curve
- **PIT**: Probability Integral Transform
- **QLIKE**: Quasi-LIKElihood (loss function)
- **QMLE**: Quasi-Maximum Likelihood Estimator
- **QQ**: Quantile–Quantile (plot)
- **QSD**: Quasi Score-Driven (model)
- **RMSE**: Root Mean Squared Error
- **RON**: Romanian leu, international currency code
- **SE**: Standard Error
- **US**: United States
- **USD**: US dollar, international currency code
- **VIX**: Chicago Board Options Exchange Volatility Index
- **XAU**: gold, international currency code (one troy ounce)

References (1/6)

- Andersen, T.G., Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4), 885–905.
- Andrews, D.W.K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68(2), 399–405.
- Andrews, D.W.K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3), 683–734.
- Baillie, R.T., Bollerslev, T., Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1), 3–30.
- Berkes, I., Horváth, L., Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli*, 9(2), 201–227.
- Blasques, F., Francq, C., Laurent, S. (2023). Quasi score-driven models. *Journal of Econometrics*, 234(1), 251–275.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
- 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.

References (2/6)

- 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.
- Christensen, K., Siggaard, M., Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, 21(5), 1680–1727.
- Clark, T.E., McCracken, M.W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105(1), 85–110.
- Conrad, C., Engle, R.F. (2025). Modelling volatility cycles: the MF2-GARCH model. *Journal of Applied Econometrics*, 40(4), 438–454.
- Creal, D., Koopman, S.J., Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5), 777–795.
- Diebold, F.X., Gunther, T.A., Tay, A.S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4), 863–883.
- Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
- Engle, R.F. (2004). Risk and volatility: econometric models and financial practice. *American Economic Review*, 94(3), 405–420.

References (3/6)

- Engle, R.F., Ghysels, E., Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, 95(3), 776–797.
- Engle, R.F., Lilien, D.M., Robins, R.P. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, 55(2), 391–407.
- Engle, R.F., Ng, V.K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5), 1749–1778.
- Fan, J., Qi, L., Xiu, D. (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics*, 32(2), 178–191.
- Francq, C., Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4), 605–637.
- Francq, C., Zakoïan, J.-M. (2019). GARCH Models: Structure, Statistical Inference and Financial Applications, 2nd ed. Wiley.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). Statistics of Financial Markets: An Introduction, 5th ed. Springer.
- Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.

References (4/6)

- 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.
- Hall, P., Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, 71(1), 285–317.
- Hamilton, J.D., Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1–2), 307–333.
- Hansen, B.E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3), 705–730.
- Hansen, P.R., Huang, Z., Shek, H.H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6), 877–906.
- 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.
- Hansen, P.R., Lunde, A., Nason, J.M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497.
- J.P. Morgan/Reuters (1996). RiskMetrics — Technical Document, 4th ed. New York.

References (5/6)

- Jensen, S.T., Rahbek, A. (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory*, 20(6), 1203–1226.
- Lamoureux, C.G., Lastrapes, W.D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, 8(2), 225–234.
- Lee, J.H.H., King, M.L. (1993). A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business & Economic Statistics*, 11(1), 17–27.
- Li, W.K., Mak, T.K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, 15(6), 627–636.
- Lumsdaine, R.L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica*, 64(3), 575–596.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36(4), 394–419.
- 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.
- Mincer, J., Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. Mincer (ed.), *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER.

References (6/6)

- Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6(3), 318–334.
- Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59(2), 347–370.
- Patton, A.J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256.
- Pele, D.T., Lazar, E., Dufour, A. (2017). Information entropy and measures of market risk. *Entropy*, 19(5), 226.
- Pele, D.T., Lazar, E., Mazurencu-Marinescu-Pele, M. (2019). Modeling expected shortfall using tail entropy. *Entropy*, 21(12), 1204.
- Pele, D.T., Mazurencu-Marinescu-Pele, M. (2019). Using high-frequency entropy to forecast Bitcoin's daily Value at Risk. *Entropy*, 21(2), 102.
- Pierce, D.A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics*, 10(2), 475–478.
- Self, S.G., Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398), 605–610.
- Sheppard, K. arch: ARCH models and other tools for financial econometrics in Python (software documentation).
- West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5), 1067–1084.