Asymmetry, Fat Tails, and the Cost of the Wrong Innovation:
A Controlled GARCH Tail-Risk Study

Eugen Soloviov

July 2026

Abstract

A daily volatility model that assumes Gaussian shocks and treats good and bad news symmetrically is cheap to fit and expensive to trust: its Value-at-Risk is optimistic in exactly the tail that matters. We quantify that cost with seeded, fully reproducible experiments on a known ground truth — a GJR-GARCH(1,1) data-generating process with standardized Student-t innovations ($\nu = 5$), a positive leverage asymmetry ($\gamma = 0.10$), and crypto-like persistence (0.98). Fitting five competing families over 120 seeds and four sample sizes, the share of runs in which BIC selects an asymmetric, heavy-tailed model rises monotonically with the sample — from 0.333 at $T = 500$ to 0.992 at $T = 4000$ — while the Engle–Ng sign-bias pre-test is a far weaker detector (power 0.058 to 0.217) than the fitted asymmetry coefficient itself (0.425 to 1.000). On the tail-risk task, a misspecified Gaussian GARCH materially under-covers: over 200 paths its one-step 99% VaR is breached at 1.58% (versus the nominal 1%) and its 99.5% VaR at 1.10%, so the Kupiec test rejects it on 80.5% and 96.5% of paths, against 4.5% for the true model. Switching only the innovation to Student-t repairs almost all of it (violation rates 1.03% and 0.52%; Expected Shortfall bias cut from -23.0% to $+0.9\%$ at 99%), and adding the asymmetric term is a second-order refinement. The tail-shape effect (Normal $\to t$) dominates the asymmetry effect (GARCH $\to$ GJR) by more than an order of magnitude in ES bias. Finally, on a near-symmetric, near-Gaussian control DGP ($\gamma = 0$, $\nu = 100$) the extra structure stops paying: BIC selects an asymmetric heavy-tailed model in at most 1.7% of runs, the sign-bias test holds its nominal size, and the richer models do not improve coverage. The deliverable is a calibrated coverage cost of the wrong innovation assumption — not a live-market VaR claim.

1 Introduction

A generalized autoregressive conditional heteroskedasticity (GARCH) model turns a history of returns into a one-step-ahead forecast of the conditional distribution, and from that distribution one reads Value-at-Risk (VaR) and Expected Shortfall (ES). The plain GARCH(1,1) of Bollerslev (1986) makes two assumptions that are convenient and, for risk work, dangerous. It is *symmetric*: a shock enters the variance recursion only through its square, so a large drop and an equally large rise raise tomorrow's variance identically. And it assumes *Gaussian* innovations, which underprice how often a return lands four or more standard deviations from the mean. Both defects push the same way at the tail: they make VaR look smaller and safer than it is.

The fixes are well established. Asymmetry is captured by the threshold term of the GJR-GARCH model (Glosten et al., 1993) or by the log-variance form of EGARCH (Nelson, 1991); the leverage story that motivates it dates to Black (1976). Fat tails are captured by Student-t innovations (Bollerslev, 1987) or, allowing a heavier left tail, Hansen's skewed-t (Hansen, 1994). Whether asymmetry is present at all can be pre-tested with the sign-bias regressions of Engle and

Ng (1993), and a fitted VaR forecast is judged by the unconditional-coverage test of Kupiec (1995) and the conditional-coverage test of Christoffersen (1998). None of this is new.

What is harder to find is a *controlled* measurement of what the wrong assumption actually costs, stated in the currency a risk desk cares about: tail-coverage error and ES bias. Real returns never reveal their true data-generating process (DGP), so on live data one cannot separate "the model is miscalibrated" from "the market did something unusual." We therefore work entirely on synthetic, seeded DGPs with *known* ground truth. We build a GJR-GARCH process with Student-t innovations whose asymmetry and tail index we choose, fit the family of competing models to it, and measure — against the truth we planted — how often the right structure is recovered and how badly the wrong structure miscovers the tail. We make no claim about any real asset; the contribution is the calibrated coverage cost and its decomposition into a tail-shape effect and an asymmetry effect, plus an honest control experiment showing when the extra structure earns nothing. This study accompanies a marketmaker.cc blog post.

Concretely, four experiments (Section 4) run against the DGP of Section 2 and the estimators of Section 3:

1. **Model selection.** Simulate the GJR-t DGP; fit five families; measure how often AIC/BIC pick an asymmetric, heavy-tailed model and how powerful the sign-bias test is, as the sample grows.

2. **VaR/ES calibration.** From a misspecified Gaussian GARCH, a GARCH-t, and the true GJR-t, produce one-step VaR/ES at three confidence levels and backtest coverage across many paths.

3. **The cost of symmetry.** Decompose the breach-rate error and ES bias into the tail-shape effect (Normal vs t) and the asymmetry effect (GARCH vs GJR).

4. **When it does not pay.** Repeat on a near-symmetric, near-Gaussian DGP and show the extra parameters stop helping — the honest counter-result.

2 The data-generating process and its ground truth

All returns are on the percent (times one hundred) scale, the scale on which GARCH optimizers are well conditioned. The primary DGP is a GJR-GARCH(1,1) with standardized Student-t innovations. Let $r_t = \mu + \varepsilon_t$ with $\varepsilon_t = \sigma_t z_t$, where z_t is drawn from the unit-variance standardized Student-t with ν degrees of freedom, and

$$\sigma_t^2 = \omega + \big(\alpha + \gamma\,\mathbf{1}\{\varepsilon_{t-1} < 0\}\big)\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2. \tag{1}$$

The threshold parameter γ is the entire asymmetry story: after a negative shock the squared-shock impact is $\alpha + \gamma$; after a non-negative shock it is α. With innovations symmetric about zero the covariance-stationary persistence is $\alpha + \beta + \frac{1}{2}\gamma$ and the unconditional variance is $\omega/(1 - \alpha - \beta - \frac{1}{2}\gamma)$.

We fix the ground truth at $\omega = 0.2$, $\alpha = 0.05$, $\gamma = 0.10$, $\beta = 0.88$, and $\nu = 5$, with μ set to zero. This gives a persistence of 0.98 — high, as daily crypto tends to be — an unconditional variance of 10 and hence a daily volatility of 3.16% (all derived from the parameters), a genuine positive leverage effect ($\gamma > 0$), and heavy but finite-variance tails ($\nu = 5$, so the fourth moment is finite). Because we generate the series, the one-step-ahead conditional standard deviation σ_t of Eq. (1) is known exactly at every date; the *true* model's VaR therefore carries no estimation error, and serves as the calibrated reference the fitted models are measured against.

The control DGP of Experiment 4 keeps the same structure but sets γ to zero (exact symmetry) and $\nu = 100$ (effectively Gaussian innovations), with $\alpha = 0.06$, $\beta = 0.90$ (persistence 0.96). Here the asymmetric, heavy-tailed features the richer models are built to exploit are absent by construction, so preferring those models is an error the selection criteria should avoid.

3 Methods

Estimation. The five competing families — symmetric GARCH-Normal, GARCH-t, GJR-Normal, GJR-t, and EGARCH-t — are fit by maximum likelihood with the `arch` package (Sheppard, 2023) (constant mean, orders $p = o = q = 1$ where the asymmetric term is present). The true family is GJR-t; we treat both GJR-t and EGARCH-t as members of the correct *class* (asymmetric and heavy-tailed), since both encode the two features the DGP possesses.

VaR and ES. For a location–scale model $r_{t+1} = \mu + \sigma_{t+1} z_{t+1}$ the one-step forecasts at confidence α are

$$\mathrm{VaR}_\alpha = -\big(\mu + \sigma_{t+1} F_z^{-1}(1-\alpha)\big), \qquad \mathrm{ES}_\alpha = -\big(\mu + \sigma_{t+1}\, \mathbb{E}[z \mid z \le F_z^{-1}(1-\alpha)]\big), \tag{2}$$

reported as positive loss numbers. F_z^{-1} is the quantile of the fitted standardized innovation. For the Normal, $\mathbb{E}[z \mid z \le q] = -\phi(q)/(1-\alpha)$; for the unit-variance standardized Student-t we use the closed-form tail expectation of McNeil et al. (2015). Because the standardized quantile and tail expectation depend only on the distribution, not on σ_{t+1}, they are computed once and applied along the whole evaluation window. To score a fitted model one-step-ahead out of sample, we freeze its estimated parameters and run the variance recursion of Eq. (1) forward over the evaluation returns, so σ_{t+1} uses only information up to t.

Sign-bias test. On the standardized residuals z_t of a fitted symmetric GARCH-Normal we run the Engle–Ng regression

$$z_t^2 = a_0 + a_1 S_{t-1}^- + a_2 S_{t-1}^- \varepsilon_{t-1} + a_3 S_{t-1}^+ \varepsilon_{t-1} + u_t, \qquad S_{t-1}^- = \mathbf{1}\{\varepsilon_{t-1} < 0\}, \tag{3}$$

and take the joint F-test of $a_1 = a_2 = a_3 = 0$ as the omnibus test for neglected asymmetry (implemented directly by least squares, no external dependency). Its power is the share of runs rejecting at the 5% level.

Backtests. With violation indicator $I_t = \mathbf{1}\{\mathrm{loss}_t > \mathrm{VaR}_t\}$, $N = \sum_t I_t$ over T days, observed rate $\hat\pi = N/T$, and target $p = 1 - \alpha$, the Kupiec unconditional-coverage statistic is

$$LR_{uc} = -2 \log \frac{p^N (1-p)^{T-N}}{\hat\pi^N (1-\hat\pi)^{T-N}} \sim \chi^2_1, \tag{4}$$

and the Christoffersen independence statistic LR_{ind} compares the first-order Markov transition likelihood of the violation sequence to its iid restriction; the conditional-coverage statistic is $LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^2_2$ (Kupiec, 1995; Christoffersen, 1998). We report, per model and level, the mean violation rate and the share of paths on which each test rejects at 5%.

Table 1: Experiment 1 (GJR-t DGP, 120 seeds per size). Share of runs in which each information criterion selects an asymmetric, heavy-tailed model (GJR-t or EGARCH-t); the share selecting the exact GJR-t family under BIC; the share in which the fitted γ is significant at 5%; and the power of the Engle–Ng sign-bias test. Correct-family recovery rises monotonically with the sample.

T	AIC asym.+heavy	BIC asym.+heavy	BIC exact GJR-t	γ signif. rate	Sign-bias power
500	0.725	0.333	0.225	0.425	0.058
1000	0.892	0.525	0.442	0.742	0.067
2000	0.992	0.875	0.808	0.967	0.092
4000	1.000	0.992	0.975	1.000	0.217

4 Experimental design

One seeded Python harness (`scripts/run_all.py`, estimators in `scripts/garch_lib.py`) generates and analyzes everything under Python 3.14.6 with NumPy 2.5.1; every random stream is seeded, so the results are bit-reproducible by one command, and `scripts/check_paper_numbers.py` verifies every numeric claim below against `results/results.json`. Quantities labeled *derived* are arithmetic on DGP parameters or on JSON entries.

Experiment 1 draws the GJR-t DGP at sample sizes $T \in \{500, 1000, 2000, 4000\}$, 120 seeds each, fits all five families, and records for every run whether the AIC- and BIC-minimizing model is asymmetric and heavy-tailed, whether it is exactly GJR-t, whether the fitted γ is significant at 5%, and whether the sign-bias test rejects. Experiments 2–3 simulate 200 independent paths of $T = 6000$ observations, fit each candidate on the first 2000 (the training window) and evaluate one-step VaR/ES coverage on the remaining 4000 (the evaluation window), at confidence levels 95%, 99%, and 99.5%. Experiment 2 reports the three headline models — the misspecified Gaussian GARCH, the GARCH-t, and the true GJR-t; Experiment 3 adds the GJR-Normal cell to complete the two-by-two decomposition of tail shape against asymmetry. Experiment 4 repeats Experiments 1–2 on the near-Gaussian control DGP.

5 Results

5.1 Model selection: recovery rises with the sample

Table 1 shows the core recovery result. At $T = 500$ the sample is too short to see the modest asymmetry cleanly: BIC, which penalizes parameters aggressively, picks an asymmetric heavy-tailed model in only 0.333 of runs, and the exact GJR-t in 0.225. As T grows the picture sharpens monotonically — BIC reaches 0.525 at $T = 1000$, 0.875 at $T = 2000$, and 0.992 at $T = 4000$. AIC, with its lighter penalty, is quicker to admit the extra structure (0.725 rising to 1.000) at the cost of occasionally over-fitting on the control DGP (Section 5.4). The fitted asymmetry coefficient is the most direct detector: the fraction of runs in which γ is significant at 5% climbs from 0.425 to 1.000.

The sign-bias test is the weak link. Its power rises with T — from 0.058 to 0.217 — but even at $T = 4000$, where the fitted γ is significant in every single run, the omnibus sign-bias pre-test rejects only about a fifth of the time. The lesson is practical: for a leverage effect of this modest size, fitting the asymmetric model and reading its γ t-statistic is far more sensitive than the residual-based pre-test. The pre-test is a conservative gate, not a substitute for fitting the richer model.

Table 2: Experiments 2 (GJR-t DGP, 200 paths, evaluation window 4000). Mean one-step VaR violation rate and the share of paths on which the Kupiec unconditional-coverage test rejects at 5%, by model and confidence level. The Gaussian GARCH under-covers the 99% and 99.5% tail and is rejected on almost every path; the Student-t models are calibrated. Nominal violation rates are 5%, 1%, 0.5%.

	Mean violation rate			Kupiec reject rate		
Model	95%	99%	99.5%	95%	99%	99.5%
True GJR-t (known)	4.99%	1.00%	0.51%	0.055	0.045	0.045
Gaussian GARCH (misspec.)	4.52%	1.58%	1.10%	0.410	0.805	0.965
GARCH-t	5.09%	1.03%	0.52%	0.225	0.190	0.145
Fitted GJR-t	4.98%	1.01%	0.52%	0.175	0.180	0.165

5.2 VaR/ES calibration: the Gaussian under-covers the tail

Table 2 is the headline. The true GJR-t model, carrying no estimation error, is calibrated by construction: violation rates of 4.99%, 1.00%, and 0.51% against nominal 5%, 1%, and 0.5%, with Kupiec rejecting on 5.5%, 4.5%, and 4.5% of paths — exactly the nominal size of the test, confirming that the backtest itself is well calibrated.

The misspecified Gaussian GARCH fails exactly where a risk model must not. At 99% its VaR is breached at 1.58% — more than half again the nominal rate — and at 99.5% at 1.10%, more than double. The Kupiec test consequently rejects it on 80.5% of paths at 99% and 96.5% at 99.5%: an analyst running this model would be told, almost every time, that it is broken. A revealing detail is that at 95% the Gaussian error changes sign — it *over*-covers, breaching at only 4.52%. A unit-variance Student-t has less mass just past its central region than a Gaussian, so the Gaussian 95% VaR is slightly too wide; only further out does the heavier t tail take over and the Gaussian become dangerously too narrow. The single-level view flatters the Gaussian precisely at the level where the tail is least interesting.

Both Student-t models are calibrated on the violation rate (1.03% and 1.01% at 99%; 0.52% at 99.5%). Their Kupiec rejection rates (0.145 to 0.225) sit above the nominal 5% but far below the Gaussian's, and the gap is the honest signature of estimation risk: unlike the true model, the fitted models freeze parameters estimated on 2000 observations and carry that estimation error into the evaluation window, which the test detects on a minority of paths. The categorical failure of the Gaussian is of a different kind entirely — a structural under-pricing of the tail, not sampling noise.

5.3 The cost of symmetry: tail shape dominates asymmetry

Table 3 decomposes the miscalibration into its two sources. Reading down the tail-shape axis (Normal to t, holding the variance structure fixed) is dramatic: at 99% the ES bias collapses from -23.0% under Gaussian GARCH to $+0.9\%$ under GARCH-t, and the breach-rate error from $+0.58$ to $+0.03$ percentage points. Reading down the asymmetry axis (GARCH to GJR, holding the innovation fixed) barely moves the numbers: Gaussian GJR is still badly biased (-22.8% ES at 99%), and GARCH-t to GJR-t shifts the ES bias only from $+0.9\%$ to essentially zero.

The magnitudes tell the story a risk manager needs. Ignoring fat tails understates 99.5% Expected Shortfall by 29.1% — nearly a third of the capital the tail actually demands — and the understatement grows with the confidence level, from 8.4% at 95% to 23.0% at 99% to 29.1% at 99.5%, because the Gaussian tail thins faster than the true t tail exactly where ES integrates. Ignoring asymmetry, by contrast, costs well under one percentage point of ES bias at every level.

Table 3: Experiment 3 (GJR-t DGP): VaR breach-rate error (observed minus nominal, in percentage points) and Expected Shortfall relative bias (percent, against the true model's ES) for the two-by-two of tail shape and asymmetry. The Normal-to-t move removes almost all of the miscalibration; the GARCH-to-GJR move is second order. ES bias grows with the confidence level.

	Breach-rate error (pp)			ES relative bias (%)		
Model	95%	99%	99.5%	95%	99%	99.5%
Gaussian GARCH	−0.48	+0.58	+0.60	−8.4	−23.0	−29.1
Gaussian GJR	−0.61	+0.52	+0.57	−7.9	−22.8	−28.9
GARCH-t	+0.09	+0.03	+0.02	+0.1	+0.9	+1.3
GJR-t	−0.02	+0.01	+0.02	0.0	0.0	0.0

Table 4: Experiment 4 (near-Gaussian control DGP, $\gamma = 0$, $\nu = 100$). Selection almost never prefers the richer class, and the sign-bias test shows no power because there is no asymmetry to find. Coverage figures at 99.5%: the Gaussian model is now calibrated too, and the richer models do not improve on it.

T	BIC asym.+heavy	AIC asym.+heavy	Sign-bias power
500	0.017	0.075	0.050
1000	0.000	0.067	0.033
2000	0.000	0.100	0.025
4000	0.000	0.075	0.000

For this DGP — and, the blog argues, for daily crypto generally — the first and by far the larger dividend is paid by the innovation distribution, not the asymmetric variance term. That does not make asymmetry worthless: the GJR-t cell is the only one whose breach rate and ES bias are both essentially zero at all three levels, and only the asymmetric model recovers the true conditional variance path. But if a modeling budget buys just one refinement, it buys the t.

5.4 When it does not pay: the honest counter-result

The features have to be present for modeling them to help. On the near-symmetric, near-Gaussian control DGP (Table 4) the selection criteria correctly refuse the extra structure: BIC picks an asymmetric heavy-tailed model in at most 1.7% of runs, falling to 0.000 for every sample size of 1000 or more, and the sign-bias test holds its nominal size (power 0.050 at $T = 500$, drifting to 0.000 at $T = 4000$ — no spurious asymmetry detected). AIC is less disciplined, admitting the richer class in up to 10.0% of runs through its lighter penalty, which is the price of AIC's known tendency to over-parameterize.

The coverage side agrees. On this DGP the Gaussian model is no longer misspecified in the tail, and its 99.5% breach-rate error is a negligible +0.05 percentage points, with an ES bias of only −1.5% — an order of magnitude smaller than the −29.1% it suffered on the true heavy-tailed DGP. Crucially, the richer models buy nothing here: the fitted GJR-t Kupiec rejection rate at 99.5% (0.075) is no better than the Gaussian's (0.115), and its extra parameters only add estimation noise. This is the symmetric statement to Sections 5.2–5.3: the extra structure pays precisely, and only, when the true process has the feature it models.

6 Discussion

Three practical conclusions follow from the controlled measurements. First, the tail-coverage cost of the wrong innovation assumption is large and one-directional: a Gaussian daily GARCH undercovers the far tail of a heavy-tailed process badly enough to be rejected on almost every path, and understates Expected Shortfall by a quarter to a third at the 99–99.5% levels that drive capital. This is not a subtle bias to be argued over; it is a structural error a coverage backtest catches immediately. Second, the fix is cheap and specific. Almost all of the miscalibration is repaired by switching the innovation from Normal to Student-t — a single extra parameter — while the asymmetric variance term, though it is what makes the DGP a GJR process, is a second-order refinement for tail coverage. An analyst forced to choose one upgrade should take the fat tails. Third, complexity must be earned: on a process without asymmetry or heavy tails, BIC and the sign-bias test correctly decline the richer models, and those models deliver no coverage improvement while adding estimation variance. The recovery experiment quantifies how much data that discipline needs — BIC does not reliably prefer the correct family until T is in the low thousands even when the feature is genuinely present.

A recurring theme is that different diagnostics have very different power. The fitted γ t-statistic detects the planted asymmetry in every run by $T = 4000$, while the Engle–Ng omnibus sign-bias pre-test, run on symmetric residuals, rejects only a fifth of the time; and BIC lags AIC in admitting real structure but is far more resistant to admitting spurious structure. The honest reading is to fit the richer model and test its parameters directly, using the residual pre-test as a conservative gate and BIC as the arbiter against over-fitting.

7 Limitations

- **Synthetic data, by design.** Both DGPs are GJR-GARCH processes with (standardized) Student-t innovations, chosen so the ground truth is known exactly. That is the point of a calibration study and also its boundary: real returns have structural breaks, intraday seasonality, and regime changes that no single stationary GARCH captures. The numbers here are the coverage cost of the innovation assumption *when the rest of the model is correct*, not a claim about any traded asset's VaR.

- **One parameterization per DGP.** The asymmetry ($\gamma = 0.10$) and tail index ($\nu = 5$) are single, crypto-motivated choices, not swept. A stronger leverage effect would raise the asymmetry dividend relative to the tail-shape dividend; a milder one would shrink it further. The qualitative ranking (tail shape first) is robust for realistic daily values but the exact magnitudes are specific to these parameters.

- **Frozen-parameter evaluation.** Fitted models are estimated once on the training window and filtered forward, rather than refit on a rolling window. This isolates the effect of the distributional assumption from that of parameter drift and keeps the experiment fast and deterministic; it also means the fitted models' mild excess Kupiec rejection is a clean measure of estimation risk at a fixed sample size, not of adaptivity.

- **One-step horizon only.** All VaR/ES forecasts are one day ahead, where the location–scale formulas of Eq. (2) are exact. Multi-day tail risk requires simulation through the recursion and is out of scope.

- **Skewed-t and EGARCH not scored for coverage.** The blog treats Hansen's skewed-t and EGARCH in depth; here EGARCH-t enters only the selection experiment and the skewed-t not at all, since the DGP innovation is symmetric-t and adding a skew term to the risk comparison would test a feature the ground truth does not contain.

8 Conclusion

On seeded GARCH processes with known ground truth, the cost of the wrong innovation assumption is measurable and blunt. A Gaussian daily GARCH fitted to a heavy-tailed, mildly asymmetric process under-covers its 99% VaR at 1.58% and its 99.5% VaR at 1.10%, is rejected by the Kupiec test on 80.5% and 96.5% of paths, and understates 99.5% Expected Shortfall by 29.1%. Switching the innovation to Student-t removes almost all of that error with one parameter; adding the asymmetric variance term is a real but second-order refinement. And when the process has neither heavy tails nor asymmetry, the selection criteria correctly stop preferring the richer models and those models stop helping. The extra structure is worth its parameters exactly when the true process carries the feature — a conclusion available only because the truth was known by construction.

Reproducibility. All code, tests, and outputs accompany this paper: `scripts/run_all.py` regenerates `results/results.json` from fixed seeds (Python 3.14.6, NumPy 2.5.1); `scripts/check_paper_numbers.py` verifies every numeric claim in this manuscript against that file and fails on any mismatch; `tests/` contains deterministic invariant tests for every estimator used.

References

Fischer Black. Studies of stock price volatility changes. In *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pages 177–181, 1976.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986. doi: 10.1016/0304-4076(86)90063-1.

Tim Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69(3):542–547, 1987. doi: 10.2307/1925546.

Peter F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4):841–862, 1998. doi: 10.2307/2527341.

Robert F. Engle and Victor K. Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5):1749–1778, 1993. doi: 10.1111/j.1540-6261.1993.tb05127.x.

Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48 (5):1779–1801, 1993. doi: 10.1111/j.1540-6261.1993.tb05128.x.

Bruce E. Hansen. Autoregressive conditional density estimation. *International Economic Review*, 35(3):705–730, 1994. doi: 10.2307/2527081.

Paul H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2):73–84, 1995. doi: 10.3905/jod.1995.407942.

Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ, 2 edition, 2015. ISBN 978-0-691-16627-8.

Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991. doi: 10.2307/2938260.

Kevin Sheppard. arch: Autoregressive conditional heteroskedasticity (arch) and other tools for financial econometrics in python. `https://github.com/bashtage/arch`, 2023. Python package for GARCH-family estimation.