Stochastic and Time-Series Foundations for GARCH

(with selected metrics implemented in `utils.py`)

Guillem Borràs

Abstract

This note provides a master's-level, conceptually focused yet mathematically rich overview of the discrete-time stochastic and time-series foundations underlying GARCH-type volatility models, and the key metrics implemented in the referenced `utils.py`. We cover returns definitions, stationarity and ergodicity, stylized facts of asset returns, ARCH/GARCH specifications and conditions, innovation distributions (Gaussian vs. Student-t), concise treatment of VaR/CVaR, diagnostics (ACF and QQ-plots), and a summary of additional performance and risk measures (Sharpe, Sortino, drawdowns, beta/alpha, *etc.*).

Contents

1 Discrete-Time Stochastic Processes and Modeling Foundations

Random walks and Brownian motion (intuition). In continuous time, a standard Brownian motion $(W_t)_{t\geq 0}$ has $W_0 = 0$, independent and stationary increments, and $W_t - W_s \sim \mathcal{N}(0, t-s)$ for $t > s$. In discrete time, the sum $S_n = \sum_{i=1}^{n} \xi_i$ with i.i.d. increments ξ_i (mean 0, finite variance) provides the canonical random-walk analogue. Many financial return models in discrete time can be viewed as stochastic difference equations for (r_t), the (log) returns.

Stochastic difference equations. A basic linear example is AR(1): $X_t = c + \phi X_{t-1} + \varepsilon_t$ with $|\phi| < 1$ and i.i.d. noise (ε_t). GARCH-type models can be seen as *difference equations for the conditional variance* (Sec. 4): σ_t^2 is driven by its lag(s) and lagged squared shocks.

Stationarity.

Definition 1 (Strict and covariance stationarity)**.** A process (Y_t) is *strictly stationary* if the joint law of $(Y_{t_1}, \dots, Y_{t_k})$ is invariant to time shifts. It is *covariance stationary* if $\mathbb{E}[Y_t] = \mu$ is constant, $\mathrm{Var}(Y_t) = \sigma^2$ is finite and constant, and $\gamma(k) = \mathrm{Cov}(Y_t, Y_{t-k})$ depends only on the lag k.

Covariance stationarity is typically assumed for (demeaned) return processes, enabling stable estimation of moments and autocovariances.

Ergodicity.

Definition 2 (Ergodicity (informal))**.** A stationary process is *ergodic* if time averages converge to ensemble averages; in practice, sample moments computed on a long realization converge (in probability) to their population counterparts.

Ergodicity justifies estimation from a single time series realization. Many ARMA and GARCH models are ergodic under standard parameter constraints.

2 Returns: Definitions, Scaling, and Aggregation

Let P_t denote the asset price at time t. The *simple* (arithmetic) return and the *log* return are

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1, \qquad r_t = \log \frac{P_t}{P_{t-1}} = \log(1 + R_t).$$

For small moves $r_t \approx R_t$, but r_t is additive across time: $\log(P_T/P_0) = \sum_{t=1}^{T} r_t$.

Annualization (daily data). If there are n_{y} trading days/year (e.g. 252): the annualized mean is $\mu_{\mathrm{ann}} \approx n_{\mathrm{y}}\, \mu_d$; the annualized volatility is $\sigma_{\mathrm{ann}} \approx \sqrt{n_{\mathrm{y}}}\, \sigma_d$.

3 Stylized Facts of Asset Returns

Common empirical regularities motivating volatility models:

- **Weak linear predictability of returns.** Daily returns r_t typically exhibit little to no autocorrelation.
- **Volatility clustering.** Large absolute returns tend to follow large absolute returns (and small follow small), implying serial dependence in $|r_t|$ and r_t^2.
- **Heavy tails.** Return distributions exhibit excess kurtosis relative to $\mathcal{N}(0, \sigma^2)$.
- **Leverage effect.** Negative returns precede higher subsequent volatility (asymmetry not captured by symmetric GARCH).
- **Persistence in volatility.** Autocorrelations of $|r_t|$ or r_t^2 decay slowly; GARCH(1,1) often captures this via a high $\alpha_1 + \beta_1$.

4 ARCH/GARCH: Conditional Heteroskedasticity Models

Mean and variance equations. Let r_t denote (demeaned) returns. The mean is often modeled as $r_t = \mu + \varepsilon_t$ with a constant μ (or ARMA, if needed). Volatility models posit

$$\varepsilon_t = \sigma_t z_t, \qquad \mathbb{E}[z_t] = 0, \quad \mathrm{Var}(z_t) = 1, \quad (z_t) \text{ i.i.d. (conditionally).}$$

An ARCH(q) model sets

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2, \qquad \alpha_0 > 0,\ \alpha_j \geq 0.$$

The widely used GARCH(p, q) extends this as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2, \qquad \omega > 0,\ \alpha_j, \beta_i \geq 0.$$

In practice, GARCH(1,1) suffices in many cases:

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

Stationarity and unconditional variance. For GARCH(1,1), covariance stationarity requires $\alpha_1 + \beta_1 < 1$. Then the unconditional (long-run) variance is

$$\mathbb{E}[\sigma_t^2] = \frac{\omega}{1 - \alpha_1 - \beta_1}.$$

The persistence of volatility shocks is governed by $\alpha_1 + \beta_1$. A convenient half-life (in days) is

$$\text{half-life} = \frac{\ln(0.5)}{\ln(\alpha_1 + \beta_1)}.$$

Estimation. Parameters $(\mu, \omega, \alpha_1, \beta_1)$ (and possibly additional noise parameters) are typically estimated by (quasi-)maximum likelihood. Model comparison can be done via log-likelihood and information criteria (AIC/BIC).

5 Innovation Distributions: Gaussian versus Student-t

The *conditional* distribution of $r_t | \mathcal{F}_{t-1}$ is often taken as Gaussian, $z_t \sim \mathcal{N}(0, 1)$. However, standardized residuals frequently retain fat tails; a Student-t innovation $z_t \sim t_\nu(0, 1)$ (variance 1) with degrees of freedom $\nu > 2$ accommodates heavier tails.

- **Gaussian GARCH.** Computationally convenient baseline; QMLE remains consistent for variance dynamics even under mild misspecification.
- **Student-t GARCH.** Adds tail parameter ν; often improves fit (higher likelihood, lower BIC) and produces more realistic tail risk forecasts.
- **Unconditional heavy tails.** Even with Gaussian innovations, the *unconditional* distribution of $\varepsilon_t = \sigma_t z_t$ is heavy-tailed due to volatility mixing; Student-t innovations further enhance tail fit at the *conditional* level.

6 Diagnostics: ACF and QQ-Plot (Model Adequacy)

After fitting, check that standardized residuals $z_t = \varepsilon_t/\sigma_t$ are approximately i.i.d. with the assumed law:

- **ACF of z_t and z_t^2.** Little to no autocorrelation should remain in z_t (mean adequacy) and in z_t^2 (volatility adequacy). Ljung–Box and Engle's ARCH tests are standard.
- **QQ-plot vs. assumed innovations.** Compare empirical quantiles of z_t to the theoretical quantiles of $\mathcal{N}(0,1)$ or t_ν. Systematic deviations in the tails indicate misspecified innovation distribution.
- **Information criteria.** Use AIC/BIC to compare Gaussian vs. Student-t (and alternative mean specifications).

7 Risk Measures (Brief): VaR and CVaR

Let returns be r (loss $L = -r$). For confidence level $q \in (0,1)$:

$$\begin{aligned}\mathrm{VaR}_q &= \inf\{x : \ \mathbb{P}(L \le x) \ge q\} = -\mathrm{Quantile}_{1-q}(r),\\ \mathrm{CVaR}_q &= \mathbb{E}[L \mid L \ge \mathrm{VaR}_q].\end{aligned}$$

Historical VaR/CVaR are computed from the empirical distribution. Under a parametric model, if $r \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathrm{VaR}_q = -\big(\mu + \sigma\, z_{1-q}\big), \qquad \mathrm{CVaR}_q = -\mu + \frac{\sigma\,\phi(z_{1-q})}{1-q},$$

where z_{1-q} is the $(1-q)$-quantile of $\mathcal{N}(0,1)$ and ϕ its pdf. With Student-t innovations, replace the Gaussian quantile/pdf accordingly.

8 Selected Metrics and Formulas Reflected in `utils.py`

Below we summarize the principal metrics and computations (notation assumes daily returns r_t and n_y trading days/year).

Annualized return and volatility. With sample mean $\hat\mu_d = \frac{1}{T}\sum_{t=1}^{T} r_t$ and std. $\hat\sigma_d$,

$$\widehat{\text{ann. return}} = n_\mathrm{y}\,\hat\mu_d, \qquad \widehat{\text{ann. vol}} = \sqrt{n_\mathrm{y}}\,\hat\sigma_d.$$

Sharpe and Sortino ratios. Let r_f be the annual risk-free rate and $r_{f,d} \approx r_f/n_\mathrm{y}$ its daily equivalent. The (annualized) Sharpe is

$$\text{Sharpe} = \frac{\widehat{\text{ann. return}} - r_f}{\widehat{\text{ann. vol}}}.$$

The Sortino replaces volatility by annualized downside deviation σ^-_ann computed using only $r_t < r_{f,d}$:

$$\text{Sortino} = \frac{\widehat{\text{ann. return}} - r_f}{\sigma^-_\mathrm{ann}}.$$

Drawdowns, Calmar ratio. Define the (gross) cumulative index $C_t = \prod_{s\le t}(1+r_s)$ with $C_0 = 1$. Drawdown $DD_t = \frac{C_t}{\max_{s\le t} C_s} - 1 \le 0$. The maximum drawdown is $\min_t DD_t$. The Calmar ratio is CAGR/|MDD|.

Hit ratio, average gain/loss, Omega. Hit ratio $= \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{r_t > 0\}$. Average gain $= \mathbb{E}[r_t \mid r_t > 0]$, average loss $= \mathbb{E}[r_t \mid r_t < 0]$. A simple Omega-like ratio used in code is

$$\Omega_0 = \frac{\sum_{t:r_t>0} r_t}{\sum_{t:r_t<0}(-r_t)},$$

interpretable as gross gains over gross losses.

Benchmark beta/alpha and correlation. Given benchmark returns $r_t^{(b)}$, beta is

$$\beta = \frac{\mathrm{Cov}(r_t, r_t^{(b)})}{\mathrm{Var}(r_t^{(b)})}.$$

With annualized means $(\mu_{\text{ann}}, \mu_{\text{ann}}^{(b)})$, the (annual) CAPM alpha is

$$\alpha_{\text{ann}} = (\mu_{\text{ann}} - r_f) - \beta\,(\mu_{\text{ann}}^{(b)} - r_f),$$

and $\mathrm{Corr}(r_t, r_t^{(b)})$ is the linear correlation.

Distributional fits (Normal vs. Student-t). *Normal MLE*: $\hat{\mu} = \bar{r},\ \hat{\sigma} = \sqrt{\frac{1}{T}\sum(r_t - \bar{r})^2}$ (MLE uses T in the denominator). *Student-t MLE*: estimate (ν, μ, σ) by maximizing $\sum_t \log f_{t_\nu}(r_t \mid \mu, \sigma)$ subject to $\nu > 2$. Model comparison via AIC/BIC.

GARCH(1,1) fit and unconditional variance. Estimate $(\mu, \omega, \alpha_1, \beta_1)$ (and ν if Student-t) by ML. If $\alpha_1 + \beta_1 < 1$, the unconditional variance is $\omega/(1 - \alpha_1 - \beta_1)$. Standardized residuals $z_t = \varepsilon_t/\sigma_t$ are used for diagnostics (ACF, QQ).

Rolling metrics. Rolling volatility over window w: $\hat{\sigma}_{t,w} = \sqrt{n_{\text{y}}} \cdot \mathrm{sd}(r_{t-w+1}, \ldots, r_t)$. Rolling Sharpe (with $r_f = 0$) is $(n_{\text{y}} \cdot \bar{r}_{t,w})/\hat{\sigma}_{t,w}$.

9 Practical Workflow Aligned with the Code

1. **Data preparation.** Clean prices (remove non-positive values, forward-fill small gaps), convert to returns (simple or log).

2. **Exploratory diagnostics.** Check histograms, skew/kurtosis, and correlations; compute summary metrics.

3. **Benchmarking.** Align series to a benchmark to compute β, α, and correlation.

4. **Volatility modeling.** Fit Normal and Student-t GARCH(1,1); compare AIC/BIC; inspect persistence $\alpha_1 + \beta_1$ and unconditional variance.

5. **Model checking.** Use ACF of z_t and z_t^2, QQ-plots, and goodness-of-fit diagnostics; refine innovation choice if needed.

6. **Risk estimation.** Produce historical and parametric VaR/CVaR; reconcile model-based tails with empirical evidence.

7. **Rolling views and reporting.** Compute rolling vol/Sharpe and generate concise commentary summarizing fit and risk characteristics.

References (suggested)

References

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

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

[3] J. D. Hamilton (1994). *Time Series Analysis*. Princeton University Press.

[4] R. S. Tsay (2010). *Analysis of Financial Time Series* (3rd ed.). Wiley.

[5] A. J. McNeil, R. Frey, and P. Embrechts (2005). *Quantitative Risk Management*. Princeton University Press.

[6] R. Cont (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1(2), 223–236.