Lecture 10 — Asymmetric Volatility Models, IGARCH, QMLE, and Diagnostic Checking

Chapter 4: from symmetric GARCH recursion to asymmetric news effects, integrated variance dynamics, and practical estimation

Jiajing Sun

School of Economics and Management, University of Chinese Academy of Sciences

Econometrics and Time Series Methods

Spring 2026

Why Lecture 10 comes right after Lecture 9

Lecture 9 introduced the basic conditional-variance language:

- volatility clustering,
- ARCH(p),
- GARCH(p, q),
- persistence and stationarity of the standard symmetric model.

Lecture 10 asks three deeper questions:

1. Why do negative and positive shocks of the same magnitude often produce different volatility responses?
2. What happens when persistence reaches the boundary and variance dynamics contain a unit root?
3. How do we estimate, test, and check ARCH/GARCH models in a mathematically careful way?

Learning goals

By the end of the lecture, students should be able to:

1. explain why symmetric GARCH models can miss the leverage effect and asymmetric news impact;
2. write down and interpret EGARCH, GJR-GARCH, and threshold-style volatility models;
3. distinguish a highly persistent GARCH from an IGARCH process and explain why IGARCH is a variance-unit-root model;
4. derive the conditional Gaussian likelihood and explain the logic of QMLE;
5. use robust covariance matrices and Wald tests in GARCH estimation;
6. perform ARCH-LM, McLeod-Li, and standardized-residual diagnostics;
7. implement a basic R workflow for fitting and comparing volatility models.

Practical plan for the three contact hours

Hour 1

Asymmetric volatility models: leverage effect, news impact curves, EGARCH, GJR-GARCH, and threshold-style specifications.

Hour 2

IGARCH and near-integrated volatility; conditional Gaussian likelihood; MLE and QMLE; robust inference and numerical issues.

Hour 3

Diagnostic checking: ARCH-LM and McLeod-Li tests; standardized residual diagnostics; R workflow for estimation, comparison, and forecasting.

Main econometric theme

We are no longer asking only whether volatility changes over time. We are asking *how* shocks change volatility, *how persistent* those changes are, and *how to estimate the variance dynamics reliably*.

Quick recap: the symmetric GARCH(1,1) benchmark

The standard benchmark from Lecture 9 is

$$\varepsilon_t = \sigma_t z_t, \qquad z_t \overset{i.i.d.}{\sim} (0,1), \qquad \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2.$$

Under the usual conditions,

$$\omega > 0, \qquad \alpha \geq 0, \qquad \beta \geq 0, \qquad \alpha + \beta < 1,$$

the process is weakly stationary and

$$\mathrm{E}(\sigma_t^2) = \mathrm{E}(\varepsilon_t^2) = \frac{\omega}{1 - \alpha - \beta}.$$

- α measures the short-run effect of new shocks.
- β measures the carry-over of past conditional variance.
- $\alpha + \beta$ is the usual persistence index.

What symmetry means in the basic GARCH model

In the symmetric GARCH model, the conditional variance depends on ε_{t-1}^2 rather than on the sign of ε_{t-1}. Thus,

$$\sigma_t^2(\varepsilon_{t-1} = a) = \sigma_t^2(\varepsilon_{t-1} = -a) \qquad \text{for every } a.$$

Implication

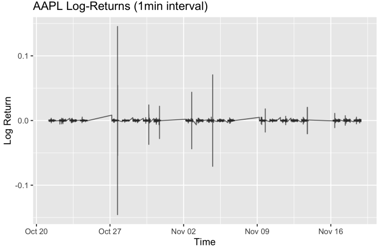
A positive surprise and a negative surprise of equal size have exactly the same effect on next period's volatility.

- This is mathematically convenient.
- But it is often too restrictive in equity and exchange-rate applications.
- In many markets, bad news raises future volatility more than good news of the same absolute size.

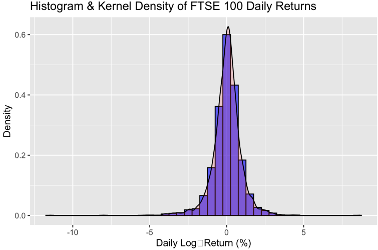
Problem

The symmetric GARCH model captures volatility clustering, but not the **asymmetric news effect** or **leverage effect**.

Stylized facts: volatility bursts and heavy-tailed returns



High-frequency returns often alternate between quiet periods and bursts of turbulence.

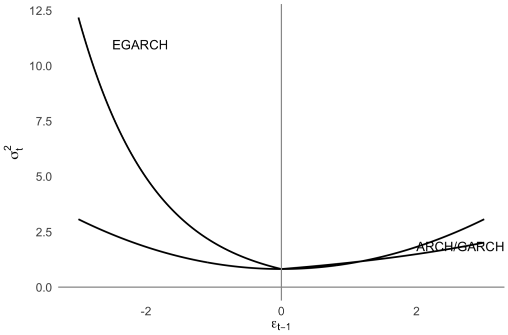


Return distributions are usually more peaked and heavier tailed than the Gaussian benchmark.

- Volatility clustering is a second-moment phenomenon.
- Heavy tails and asymmetry in the response to bad news are distinct issues.
- A good empirical model often needs both dynamic variance recursion and a non-Gaussian conditional distribution.

The leverage effect and the news impact curve

A convenient way to visualize asymmetry is the **news impact curve**: the graph of next period's conditional variance as a function of today's shock, keeping lagged variance fixed.



- For a symmetric GARCH model, the curve is symmetric around zero.
- For an asymmetric model, the left-hand side and right-hand side differ.
- In equity markets, negative returns typically generate a steeper increase in volatility.

EGARCH(1,1): exponential GARCH

Nelson's EGARCH model writes the dynamics in log variance:

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \alpha \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}}.$$

Often we define the standardized innovation

$$z_{t-1} = \frac{\varepsilon_{t-1}}{\sigma_{t-1}},$$

so the recursion becomes

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \gamma z_{t-1} + \alpha |z_{t-1}|.$$

Interpretation

The $|z_{t-1}|$ term measures the size effect, while the z_{t-1} term measures the sign effect.

Why EGARCH is attractive

The EGARCH specification has three appealing features:

1. Because the model is written for $\log \sigma_t^2$, positivity is automatic; there is no need to impose nonnegativity on all coefficients.
2. Asymmetry appears directly through the sign term γz_{t-1}.
3. Large standardized shocks can have a nonlinear effect on volatility.

Economic interpretation of γ

If $\gamma < 0$, then negative shocks increase future volatility more than positive shocks of the same size.

- This is a natural way to model leverage in stock returns.
- The persistence parameter is tied to β in the log-variance dynamics.

A useful benchmark: the EGARCH news impact curve

Holding σ_{t-1}^2 fixed, the EGARCH(1,1) curve is driven by

$$\log \sigma_t^2 = \text{constant} + \gamma z_{t-1} + \alpha |z_{t-1}|.$$

Therefore,

$$\frac{\partial \log \sigma_t^2}{\partial z_{t-1}} = \gamma + \alpha \, \text{sgn}(z_{t-1}).$$

- The slope differs on the negative and positive sides.
- For $z_{t-1} < 0$, the slope is $\gamma - \alpha$.
- For $z_{t-1} > 0$, the slope is $\gamma + \alpha$.

Key point

The EGARCH model makes asymmetry a primitive parameterization of the volatility recursion rather than an incidental empirical outcome.

GJR-GARCH(1,1)

Another common asymmetric model is the GJR-GARCH(1,1):

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

- The indicator turns on an extra variance effect when the lagged shock is negative.
- If $\gamma > 0$, bad news has a larger effect than good news.
- The symmetric GARCH model is recovered when $\gamma = 0$.

Positivity restrictions

A sufficient set of conditions is

$$\omega > 0, \qquad \alpha \geq 0, \qquad \alpha + \gamma \geq 0, \qquad \beta \geq 0.$$

Persistence and news impact in GJR-GARCH

Suppose the innovation distribution is symmetric so that

$$\mathrm{P}(\varepsilon_{t-1} < 0) = \frac{1}{2}.$$

Then the expected coefficient on lagged squared shocks is

$$\alpha + \frac{\gamma}{2}.$$

Hence an intuitive persistence index is

$$\alpha + \beta + \frac{\gamma}{2}.$$

Why this matters

A model can display stronger apparent persistence simply because negative shocks activate the threshold effect frequently.

- The news impact curve is piecewise quadratic.
- The left branch is steeper than the right branch when $\gamma > 0$.

Threshold GARCH / TARCH intuition

A threshold-style model can also be written in standard deviation rather than variance, for example

$$\sigma_t = \psi + \beta\sigma_{t-1} + \alpha_1|\varepsilon_{t-1}|1(\varepsilon_{t-1} < 0) + \alpha_2|\varepsilon_{t-1}|1(\varepsilon_{t-1} \geq 0).$$

- This makes the asymmetry very transparent.
- If $\alpha_1 > \alpha_2$, negative shocks produce a larger increase in volatility than positive shocks.
- Threshold formulations are often easy to interpret graphically through the news impact curve.

Takeaway

There is not one unique asymmetric volatility model. Different parameterizations trade off interpretability, positivity, and ease of estimation.

Comparing symmetric and asymmetric volatility specifications

Model	Recursion scale	Asymmetry mechanism	Practical advantage
GARCH	σ_t^2	none	simple benchmark
EGARCH	$\log \sigma_t^2$	sign term γz_{t-1}	positivity automatic
GJR-GARCH	σ_t^2	threshold indicator	easy leverage interpretation
TARCH / TGARCH	σ_t or σ_t^2	separate positive/negative coefficients	direct news-impact interpretation

- No single specification dominates in every application.
- Model choice is empirical as well as theoretical.
- This is exactly why likelihood comparison and diagnostic checking matter.

How to read the asymmetric coefficients empirically

Suppose we estimate one of the asymmetric models:

- In EGARCH, $\widehat{\gamma} < 0$ suggests a leverage effect.
- In GJR-GARCH, $\widehat{\gamma} > 0$ suggests bad news raises volatility more strongly.
- If the sign coefficient is tiny and insignificant, the symmetric GARCH benchmark may be adequate.

But significance is not enough

Even when the asymmetry coefficient is statistically significant, the fitted model should still be checked using standardized residuals, residual squares, and model-comparison criteria.

Econometric warning

A large asymmetric coefficient can also be compensating for misspecified mean dynamics, outliers, heavy tails, or structural breaks in volatility.

From persistence to integration in variance

In the symmetric GARCH(1,1) model,

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

weak stationarity requires

$$\alpha + \beta < 1.$$

The closer $\alpha + \beta$ is to one, the more persistent volatility becomes.

Question

What happens exactly at the boundary $\alpha + \beta = 1$?

- This boundary case is not just "very persistent" GARCH.
- It is the variance analogue of a unit-root process.
- That is why the model is called integrated GARCH, or IGARCH.

Weak stationarity revisited

Assuming $\mathrm{E}(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ and weak stationarity,

$$\mathrm{E}(\varepsilon_t^2) = \mathrm{E}(\sigma_t^2) = \sigma^2.$$

Taking expectations in the GARCH(1,1) recursion gives

$$\sigma^2 = \omega + (\alpha + \beta)\sigma^2,$$

so

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}.$$

Boundary logic

If $\alpha + \beta \uparrow 1$, the unconditional variance explodes. At $\alpha + \beta = 1$, the weakly stationary variance formula no longer exists.

Definition of IGARCH(1,1)

The IGARCH(1,1) model is the GARCH(1,1) model with

$$\alpha + \beta = 1.$$

Then

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + (1-\alpha)\sigma_{t-1}^2.$$

- Past shocks to variance do not die out geometrically in the usual way.
- The process behaves like a unit-root model in second moments.
- Volatility shocks become permanent in the same sense that level shocks are permanent in a random walk.

Economic interpretation

IGARCH says that the market keeps a permanent memory of past volatility innovations.

Variance-unit-root representation

Define the volatility innovation

$$u_t = \frac{\varepsilon_t}{\sigma_t}, \qquad u_t^2 - 1 \text{ is mean zero conditional on } \mathcal{F}_{t-1},$$

and define

$$u_t^v = \varepsilon_t^2 - \sigma_t^2.$$

Under IGARCH,

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 - \beta v_{t-1}^v.$$

This has the same flavor as a unit-root recursion:

$$(1 - L)\sigma_t^2 = \omega - \beta v_{t-1}^v.$$

Analogy

ARIMA puts a unit root in the conditional mean. IGARCH puts a unit root in the conditional variance.

Recursive solution and conditional expectation under IGARCH

By repeated substitution,

$$\sigma_t^2 = \omega t + \sigma_0^2 - \beta \sum_{\ell=1}^{t-1} \nu_{t-\ell}^{v}.$$

Conditioning on the initial information set,

$$\mathrm{E}(\sigma_t^2 \mid \mathcal{F}_0) = \omega t + \sigma_0^2.$$

- If $\omega > 0$, the conditional mean of variance grows linearly with time.
- Therefore the process has no finite, time-invariant unconditional variance.

Important difference from stationary GARCH

In ordinary GARCH, the variance reverts toward a finite unconditional level. In IGARCH, there is no such finite long-run level in the weakly stationary sense.

ARMA representation for squared returns

Recall the general GARCH(1,1) identity

$$\varepsilon_t^2 = \omega + (\alpha + \beta)\varepsilon_{t-1}^2 - \beta \nu_{t-1}^v + \nu_t^v.$$

Under IGARCH, $\alpha + \beta = 1$, so

$$\varepsilon_t^2 = \omega + \varepsilon_{t-1}^2 - \beta \nu_{t-1}^v + \nu_t^v.$$

Equivalently,

$$(1 - L)\varepsilon_t^2 = \omega + (1 - \beta L)\nu_t^v.$$

Interpretation

The squared series itself contains a unit root, so shocks to the variance proxy ε_t^2 never disappear asymptotically.

RiskMetrics and EWMA as a special IGARCH case

When $\omega = 0$, the IGARCH recursion becomes the exponentially weighted moving average (EWMA):

$$\sigma_t^2 = \beta\sigma_{t-1}^2 + (1-\beta)\varepsilon_{t-1}^2, \qquad 0 < \beta < 1.$$

Equivalently,

$$\sigma_t^2 = (1-\beta)\sum_{j=1}^{\infty}\beta^{j-1}\varepsilon_{t-j}^2.$$

- This is the RiskMetrics variance filter.
- The parameter β is the decay factor.
- For daily data, the empirical rule of thumb $\beta = 0.94$ is widely used.

Important nuance

EWMA is operationally useful for risk measurement, even though it sits exactly on the persistence boundary.

Near-IGARCH estimates in practice

Empirical GARCH fits often produce

$$\widehat{\alpha} + \widehat{\beta} \approx 1.$$

How should we interpret that?

1. It may reflect genuinely very persistent volatility.
2. It may reflect omitted asymmetry or structural breaks.
3. It may reflect low-frequency changes in the volatility intercept that a stationary GARCH is trying to mimic.

Practical lesson

A near-unit-root GARCH estimate should not automatically be read as proof that volatility shocks are truly permanent.

Mean dynamics first, variance dynamics second

In applications we usually start from

$$y_t = \mu_t + \varepsilon_t, \qquad \mu_t = \mathrm{E}(y_t \mid \mathcal{F}_{t-1}), \qquad \varepsilon_t = \sigma_t z_t.$$

Then we specify a variance recursion such as

$$\sigma_t^2(\theta) = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2(\theta)$$

or one of the asymmetric alternatives.

Identification issue

If the conditional mean is misspecified, residual-based evidence of ARCH effects may be contaminated by mean misspecification.

So in practice we model the mean and variance jointly, even if the conditional mean is very simple for daily returns.

Conditional Gaussian log-likelihood

For illustration, consider GARCH(1,1) with parameter vector

$$\theta = (\omega, \alpha, \beta)'.$$

Under a conditional Gaussian assumption,

$$\ell_T(\theta) = \sum_{t=2}^{T} \ell_t(\theta), \qquad \ell_t(\theta) = -\frac{1}{2}\log \sigma_t^2(\theta) - \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2(\theta)}.$$

The MLE is

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \ell_T(\theta).$$

- The recursion $\sigma_t^2(\theta)$ must be evaluated for every trial parameter vector.
- This is why GARCH estimation is inherently nonlinear and numerical.

Initialization of the variance recursion

To evaluate the likelihood, we need a starting value for $\sigma_1^2(\theta)$. Common choices include:

1. the stationary mean, $\sigma_1^2(\theta) = \omega/(1-\alpha-\beta)$, when stationarity is imposed;
2. the sample average of squared residuals, $T^{-1}\sum_{t=1}^{T}\varepsilon_t^2$;
3. the first squared residual, ε_1^2;
4. joint estimation of the initial variance as an additional parameter.

Asymptotic idea

Under standard regularity conditions, the exact initialization matters less and less as T grows, but in small samples it can affect optimization and finite-sample estimates.

Score recursion

Let

$$\tilde{\varepsilon}_t^2(\theta) = \frac{\varepsilon_t^2}{\sigma_t^2(\theta)}.$$

Then the score contribution can be written as

$$\frac{\partial \ell_t}{\partial \theta}(\theta) = -\frac{1}{2}\left(\tilde{\varepsilon}_t^2(\theta) - 1\right)\frac{\partial \log \sigma_t^2(\theta)}{\partial \theta}.$$

For GARCH(1,1),

$$\frac{\partial \sigma_t^2}{\partial \omega} = 1 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \omega},$$

$$\frac{\partial \sigma_t^2}{\partial \alpha} = \varepsilon_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \alpha}, \qquad \frac{\partial \sigma_t^2}{\partial \beta} = \sigma_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2}{\partial \beta}.$$

Message

The likelihood is recursive, and the gradient is recursive too.

Hessian and information matrices

The Hessian contribution is

$$\frac{\partial^2 \ell_t}{\partial\theta\partial\theta'} = -\frac{1}{2}(\tilde{\varepsilon}_t^2 - 1)\frac{\partial^2 \log\sigma_t^2}{\partial\theta\partial\theta'} + \frac{1}{2}\tilde{\varepsilon}_t^2 \frac{\partial \log\sigma_t^2}{\partial\theta}\frac{\partial \log\sigma_t^2}{\partial\theta'}.$$

Define

$$\mathcal{I} = \mathrm{E}\left(\frac{\partial \ell_t}{\partial\theta}\frac{\partial \ell_t}{\partial\theta'}\right), \qquad \mathcal{J} = -\mathrm{E}\left(\frac{\partial^2 \ell_t}{\partial\theta\partial\theta'}\right).$$

- Under correct Gaussian specification, $\mathcal{I} = \mathcal{J}$.
- Under QMLE, the two matrices need not coincide.

MLE versus QMLE

If the conditional distribution is correctly specified as Gaussian, the estimator is the ordinary conditional MLE.
If the Gaussian likelihood is used only as a convenient objective function while the true conditional distribution is not Gaussian, the estimator is the **quasi-maximum likelihood estimator** (QMLE).

Why QMLE is important

For financial returns, the Gaussian assumption is often false, yet the conditional mean and variance may still be correctly specified. Then Gaussian QMLE can remain consistent and asymptotically normal.

- "Quasi" does *not* mean ad hoc.
- It means we rely on correct first and second conditional moments, not on a fully correct density.

QMLE asymptotic distribution

Under standard conditions for semi-strong GARCH models,

$$T^{1/2}(\widehat{\theta} - \theta_0) \Longrightarrow N\left(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}\right).$$

Here

$$\mathcal{J} = \mathrm{E}\left(-\frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta'}\right), \qquad \mathcal{I} = \mathrm{E}\left(\frac{\partial \ell_t(\theta_0)}{\partial\theta}\frac{\partial \ell_t(\theta_0)}{\partial\theta'}\right).$$

Sandwich form

The matrix $\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$ is the analogue of the robust covariance formula in nonlinear regression and GMM.

Special cases of the asymptotic covariance

If z_t is i.i.d. and Gaussian, then

$$\mathrm{E}[(z_t^2 - 1)^2] = 2, \qquad \mathcal{I} = \mathcal{J},$$

so the usual inverse-information formula is valid.
If z_t is i.i.d. but not Gaussian, then one obtains a scale adjustment:

$$\mathrm{Var}_\infty(\widehat{\theta}) = \mathrm{E}[(z_t^2 - 1)^2] \, \mathcal{J}^{-1}.$$

More generally, with only martingale-difference assumptions,

$$\widehat{V}_{\mathrm{rob}} = \widehat{\mathcal{J}}^{-1} \widehat{\mathcal{I}} \widehat{\mathcal{J}}^{-1}$$

is a consistent covariance estimator.

Practical implication

In empirical work, robust standard errors should be reported routinely for GARCH estimates.

Wald tests and linear restrictions

Suppose we want to test

$$H_0 : R\theta = r.$$

Then the robust Wald statistic is

$$W = T(R\widehat{\theta} - r)' \left(R\widehat{\mathcal{J}}^{-1}\widehat{\mathcal{I}}\widehat{\mathcal{J}}^{-1}R' \right)^{-1} (R\widehat{\theta} - r),$$

which is asymptotically χ^2_q when R has q rows.
Examples:

- $H_0 : \gamma = 0$ in EGARCH or GJR-GARCH tests symmetry.
- $H_0 : \alpha + \beta = 1$ tests an IGARCH-type boundary restriction.
- Joint restrictions can compare nested models.

Non-Gaussian conditional likelihoods

The Gaussian likelihood is not the only choice. Common alternatives include:

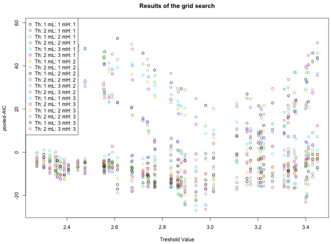
- Student-t conditional densities,
- skewed Student-t densities,
- generalized error distributions.

Why they matter

If standardized returns are heavy tailed, a Student-t likelihood can improve fit, reduce sensitivity to outliers, and deliver better volatility forecasts.

- The variance recursion may stay the same.
- Only the conditional density of z_t changes.
- So we must separate "dynamic variance misspecification" from "distributional misspecification."

Optimization is not trivial in GARCH estimation



- The likelihood can be flat near the persistence boundary.
- Different starting values may lead to different local solutions.
- Constraints such as positivity and stationarity must be enforced or reparameterized.
- Numerical convergence should always be checked; a printed coefficient table is not enough.

Practical lesson

Model estimation is an optimization problem layered on top of a recursive stochastic model. Both layers can fail.

Why diagnostics matter before and after estimation

There are two distinct diagnostic stages.

1. **Before** fitting a volatility model: test whether the residuals of the conditional mean equation show ARCH effects at all.
2. **After** fitting a volatility model: check whether the standardized residuals behave like an innovation sequence with no remaining linear or second-moment dependence.

If we skip the first stage

We may fit GARCH to noise that does not actually require a conditional variance model.

If we skip the second stage

We may report a fitted GARCH model that still leaves systematic structure in z_t or z_t^2.

ARCH-LM logic

Suppose the conditional mean model is

$$Y_t = \mu(\mathcal{F}_{t-1}, \theta_0) + \varepsilon_t, \qquad \mathrm{E}(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0.$$

Under the null of conditional homoskedasticity,

$$H_0 : \mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1}) = \sigma_0^2.$$

The alternative is that

$$\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1})$$

changes with past information.

Basic idea

If volatility is clustered, then squared residuals should be predictable from their own lags.

ARCH-LM auxiliary regression and test statistic

Estimate the mean model, obtain residuals $\widehat{\varepsilon}_t$, and run the auxiliary regression

$$\widehat{\varepsilon}_t^2 = \phi_0 + \phi_1 \widehat{\varepsilon}_{t-1}^2 + \cdots + \phi_p \widehat{\varepsilon}_{t-p}^2 + u_t.$$

Under

$$H_0 : \phi_1 = \cdots = \phi_p = 0,$$

there are no ARCH effects up to lag p.

LM statistic

Using the R^2 from the auxiliary regression,

$$LM = TR^2 \Longrightarrow \chi_p^2.$$

- This is Engle's ARCH-LM test.
- Mean-equation estimation error does not affect the asymptotic null distribution under regularity conditions.

McLeod-Li / portmanteau test for squared residual autocorrelation

A Box-Pierce-type alternative uses the sample autocorrelations of the squared residuals:

$$\widehat{\rho}_2(j) = \frac{\widehat{\gamma}_2(j)}{\widehat{\gamma}_2(0)},$$

where $\widehat{\gamma}_2(j)$ is the sample autocovariance of $\widehat{\varepsilon}_t^2$ at lag j. Then the McLeod-Li statistic is

$$ML(p) = T \sum_{j=1}^{p} \widehat{\rho}_2^2(j).$$

Under the null of no ARCH effects,

$$ML(p) \Longrightarrow \chi_p^2.$$

Interpretation

ARCH-LM and McLeod-Li are asymptotically equivalent ways to test whether the squared residuals are serially dependent.

After fitting: standardized residuals

Once a volatility model is fitted, define the standardized residuals

$$\widehat{z}_t = \frac{\widehat{\varepsilon}_t}{\widehat{\sigma}_t}.$$

A satisfactory fit should leave:

- little or no serial correlation in $\widehat{z}_t$,
- little or no serial correlation in $\widehat{z}_t^2$,
- no strong remaining ARCH-LM evidence,
- distributional diagnostics that are at least consistent with the chosen conditional density.

Why standardization matters

Raw residuals inherit time-varying variance mechanically. Standardized residuals are the object that should behave like the innovation sequence.

What a good fitted volatility model should leave behind

Check	Desired outcome
ACF of $\widehat{z}_t$	no remaining linear dependence
ACF of $\widehat{z}_t^2$	no remaining volatility clustering
Ljung-Box on $\widehat{z}_t$	insignificant or economically small
Ljung-Box on $\widehat{z}_t^2$	insignificant or economically small
ARCH-LM on $\widehat{z}_t$	fail to reject no remaining ARCH
Sign-bias / asymmetry tests	no leftover asymmetric news effect
Information criteria / log-likelihood	competitive among candidate models

Model comparison is not a one-number exercise

When comparing GARCH-family models, do not rely only on one output line.

1. Compare log-likelihood and information criteria.
2. Inspect robust standard errors and parameter significance.
3. Check whether the asymmetry coefficient is economically interpretable.
4. Check residual and squared-residual diagnostics.
5. Compare out-of-sample variance forecasts if forecasting is the goal.

Typical empirical pattern

An asymmetric model may improve the likelihood but still leave residual ARCH if the density is wrong, or vice versa.

R workflow: from data to a volatility model

A practical workflow is:

1. obtain returns and clean the series;
2. choose a simple conditional mean model;
3. test for ARCH effects in the residuals;
4. fit several candidate variance models;
5. compare them using likelihood and diagnostics;
6. produce conditional-volatility and forecast outputs.

Important

The variance model is not estimated in a vacuum. It sits on top of the data choice, frequency choice, mean specification, and assumed conditional distribution.

R code: data and conditional mean specification

```
library(quantmod)
library(forecast)
library(rugarch)
library(FinTS)

spy_xts <- getSymbols("SPY", src = "yahoo",
                    from = "2024-01-01",
                    auto.assign = FALSE)
spy_ret <- na.omit(diff(log(Ad(spy_xts))))
r <- as.numeric(spy_ret)

mean_fit <- auto.arima(r,
                     seasonal = FALSE,
                     stepwise = FALSE,
                     approximation = FALSE)
res <- residuals(mean_fit)
ArchTest(res, lags = 10)
```

- First model the mean, even if the mean dynamics are weak.
- Then test the residuals for conditional heteroskedasticity.

R code: fit symmetric and asymmetric GARCH models

```
spec_s <- ugarchspec(
  variance.model = list(model = "sGARCH", garchOrder = c(1,1)),
  mean.model = list(armaOrder = c(0,0), include.mean = TRUE),
  distribution.model = "std")

spec_e <- ugarchspec(
  variance.model = list(model = "eGARCH", garchOrder = c(1,1)),
  mean.model = list(armaOrder = c(0,0), include.mean = TRUE),
  distribution.model = "std")

spec_gjr <- ugarchspec(
  variance.model = list(model = "gjrGARCH", garchOrder = c(1,1)),
  mean.model = list(armaOrder = c(0,0), include.mean = TRUE),
  distribution.model = "std")

fit_s <- ugarchfit(spec_s, data = r)
fit_e <- ugarchfit(spec_e, data = r)
fit_gjr <- ugarchfit(spec_gjr, data = r)
```

R code: diagnostics and forecasts

```
infocriteria(fit_s)
infocriteria(fit_e)
infocriteria(fit_gjr)

z_s <- residuals(fit_s, standardize = TRUE)
Box.test(z_s, lag = 20, type = "Ljung")
Box.test(z_s^2, lag = 20, type = "Ljung")
ArchTest(as.numeric(z_s), lags = 10)

fc_s <- ugarchforecast(fit_s, n.ahead = 20)
fc_e <- ugarchforecast(fit_e, n.ahead = 20)
fc_g <- ugarchforecast(fit_gjr, n.ahead = 20)

sigma(fc_g)
```

- Compare in-sample fit and residual diagnostics.
- Then compare forecasted conditional standard deviations or variances.

How to read a fitted GARCH output economically

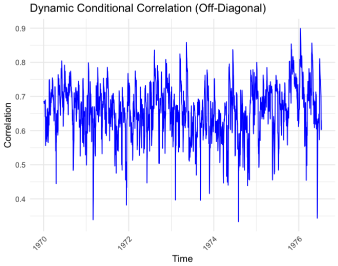
Suppose a fitted GARCH or GJR-GARCH report shows:

- a significant α — volatility reacts to new information quickly;
- a large β — volatility is persistent;
- a significant threshold parameter — bad news matters more than good news;
- Student-t degrees of freedom not too large — heavy tails remain important after volatility standardization.

Interpretation strategy

Translate coefficient estimates into persistence, asymmetry, mean reversion, and forecast implications. Do not stop at statistical significance.

Why this lecture naturally leads to multivariate volatility



Lecture 10 is still univariate. But in practice we often need time-varying covariances as well as time-varying variances.

- portfolio risk,
- contagion,
- hedge ratios,
- dynamic correlations.

This is the bridge to the multivariate-volatility material in the next lecture block.

Takeaways

- Symmetric GARCH captures clustering but not asymmetric news effects.
- EGARCH, GJR-GARCH, and threshold models encode leverage in different ways.
- IGARCH is a variance-unit-root model: volatility shocks can persist indefinitely.
- Conditional Gaussian likelihood leads naturally to MLE and QMLE.
- ARCH-LM, McLeod-Li, and standardized-residual diagnostics remain central to model checking.
- Empirical work should compare several candidate models instead of trusting one fit.

Preview

Lecture 11 moves from univariate volatility to covariance dynamics, EWMA matrices, and compact MGARCH parameterizations.