Lecture 9 — Structural VAR Identification, Volatility Clustering, ARCH, and GARCH Models

Transition from Chapter 3 to Chapter 4: from dynamic causal interpretation to dynamic conditional variance

Jiajing Sun

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

Econometrics and Time Series Methods
Spring 2026

Why this lecture bridges Chapters 3 and 4

Lecture 8 finished the **nonstationary mean-dynamics** block:

- nonstationary VAR systems,
- cointegration,
- VECMs and long-run equilibrium.

Lecture 9 completes the multivariate linear chapter by asking two new questions:

1. How do we identify economically meaningful **structural shocks** rather than merely reduced-form innovations?
2. Even if the mean equation is correct, why do financial series still show bursts of volatility and quiet periods?

Big transition

The first hour is about **structural interpretation of shocks**. The next two hours are about **time-varying conditional variance**.

Learning goals

By the end of Lecture 9, students should be able to:

1. distinguish reduced-form and structural VARs;
2. explain why identification restrictions are necessary in an SVAR;
3. describe short-run, long-run, and A/B/AB identification schemes;
4. understand volatility clustering as dependence in second moments rather than first moments;
5. write down ARCH(p) and GARCH(p, q) models and interpret their parameters;
6. explain persistence, weak stationarity, and the difference between weak and strict stationarity in GARCH systems.

Practical plan for the three contact hours

Hour 1

Structural VARs: reduced-form versus structural shocks; short-run recursive restrictions; long-run restrictions; A-, B-, and AB-models.

Hour 2

Volatility clustering; white noise versus conditional heteroskedasticity; the ARCH model; stationarity of ARCH.

Hour 3

The GARCH model; persistence and ARCH(∞) interpretation; weak stationarity and Nelson's strict-stationarity condition; bridge to Lecture 10.

Why reduced-form VARs are not enough for causal interpretation

Lectures 6–8 used reduced-form systems of the form

$$y_t = c + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t, \qquad \mathrm{E}(\varepsilon_t \varepsilon_t') = \Omega.$$

- These models describe dynamic dependence very well.
- But the innovation vector ε_t is only a mixture of underlying economic shocks.
- If we want to ask what happens after a *monetary policy shock*, a *supply shock*, or a *demand shock*, the reduced form is not enough.

Need for structure

We must separate **economically meaningful structural shocks** from **statistical reduced-form innovations**.

Structural VAR(p)

A structural VAR can be written as

$$A_0 \mathrm{y}_t = \mathrm{c}^* + A_1 \mathrm{y}_{t-1} + \cdots + A_p \mathrm{y}_{t-p} + \mathrm{u}_t,$$

where:

- A_0 captures **contemporaneous interactions**,
- $A_1, \ldots, A_p$ capture lagged effects,
- u_t is the vector of **structural shocks**.

Typically,

$$\mathrm{E}(\mathrm{u}_t) = 0, \qquad \mathrm{E}(\mathrm{u}_t \mathrm{u}_t') = \Sigma_u,$$

with Σ_u often normalized to be diagonal or even the identity.

From the structural form to the reduced form

Premultiply by A_0^{-1}:

$$\mathrm{y}_t = A_0^{-1}\mathrm{c}^* + A_0^{-1}A_1\mathrm{y}_{t-1} + \cdots + A_0^{-1}A_p\mathrm{y}_{t-p} + A_0^{-1}\mathrm{u}_t.$$

Define

$$\mathrm{c} = A_0^{-1}\mathrm{c}^*, \qquad \Phi_j = A_0^{-1}A_j, \qquad \varepsilon_t = A_0^{-1}\mathrm{u}_t.$$

Then the reduced form is

$$\mathrm{y}_t = \mathrm{c} + \Phi_1\mathrm{y}_{t-1} + \cdots + \Phi_p\mathrm{y}_{t-p} + \varepsilon_t.$$

Key covariance identity

$$\Omega = \mathrm{E}(\varepsilon_t\varepsilon_t') = A_0^{-1}\Sigma_u A_0^{-1\prime}.$$

Why the reduced-form shocks are mixtures

Even if structural shocks are economically distinct, the reduced-form innovations satisfy

$$\varepsilon_t = A_0^{-1}\mathrm{u}_t.$$

- If A_0 is not diagonal, each element of ε_t is a linear combination of multiple structural shocks.
- Therefore, a one-unit reduced-form shock does not correspond to one clean economic experiment.
- Reduced-form IRFs answer "what happens after an innovation in this equation," not necessarily "what happens after a structural policy shock."

SVAR objective

Find enough economically justified restrictions to recover u_t from ε_t.

Identification counting: why restrictions are unavoidable

The structural model contains

$$n + (p+1)n^2 + \frac{n(n+1)}{2}$$

unknowns: intercept, $(p+1)$ coefficient matrices, and the symmetric covariance matrix of structural shocks.

The reduced form contains only

$$n + pn^2 + \frac{n(n+1)}{2}$$

identified objects.

Gap

We are short of exactly n^2 restrictions.

If we normalize $\Sigma_u = I_n$, that already imposes $n(n+1)/2$ restrictions, but we still need

$$n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$$

additional restrictions.

Order condition versus rank condition

Order condition

The model must impose at least as many restrictions as there are unidentified parameters.

Rank condition

Those restrictions must also be *informative*; they cannot be algebraically redundant.

What this means in practice

It is not enough to "write down some zeros." The zeros must actually identify the structural parameters uniquely, at least locally.

Short-run identification: recursive systems

A popular identification strategy imposes contemporaneous zero restrictions on A_0. The most common case is a recursive, or triangular, system:

$$A_0 = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ a_{n1} & \cdots & a_{n,n-1} & a_{nn} \end{bmatrix}.$$

Often the diagonal is normalized to one. Then

- variable 1 does not react contemporaneously to the others,
- variable 2 reacts contemporaneously only to variable 1,
- and so on.

Econometric meaning

The ordering of the variables becomes an identifying assumption.

Why Cholesky-type identification works

If $\Sigma_u = I_n$ and A_0 is lower triangular, then the number of free off-diagonal elements is

$$\frac{n(n-1)}{2},$$

exactly matching the remaining number of restrictions needed.

Interpretation

A recursive SVAR is essentially a **Cholesky orthogonalization** with an economic ordering attached to it.

Strength and weakness

It is simple and widely used, but the results can change if we reorder the variables. So the ordering must be defended economically.

A simple macro recursive ordering

The textbook gives a monetary example with

$$y_t = \begin{bmatrix} p_t \\ gdp_t \\ m_t \\ i_t \end{bmatrix},$$

and lower-triangular A_0.

One interpretation is:

- prices do not react within the period to output, money, or the interest rate;
- output reacts within the period to prices, but not to money or the interest rate;
- money reacts within the period to prices and output;
- the policy rate can react contemporaneously to all variables.

This is an economic story, not a theorem

Different institutions or data frequencies can justify different contemporaneous restrictions.

Long-run identification

Another strategy identifies structural shocks through long-run restrictions. Suppose the structural VMA exists:

$$y_t = \sum_{j=0}^{\infty} D_j u_{t-j} = D(L) u_t.$$

Then the long-run impact matrix is

$$D(1) = \sum_{j=0}^{\infty} D_j.$$

Long-run zero restriction

If shock j has no permanent effect on variable i, impose

$$D_{ij}(1) = 0.$$

Interpretation

Short-run movements are allowed, but the cumulative long-run response must vanish.

Economic logic of long-run restrictions

Long-run restrictions are attractive when theory speaks more clearly about *permanent* effects than about *within-period* timing.

Examples:

- a demand shock may be allowed to move output in the short run but not in the long run;
- a supply or technology shock may be allowed to shift long-run output.

Practical contrast

- Short-run restrictions say who can move contemporaneously.
- Long-run restrictions say who can move the permanent component.

Other identification approaches

The textbook briefly notes several other strategies:

- **sign restrictions**: impose only the signs of selected responses;
- **identification from heteroskedasticity**: exploit variance changes across regimes;
- **external instruments / proxy SVARs**: use outside shocks or narrative events;
- **narrative restrictions**: use historically identified episodes.

Big picture

There is no single universal SVAR. Identification is a substantive modelling decision, not a purely statistical default.

A-model

In the A-model, the structural form is written as

$$A\mathrm{y}_t = \mathrm{c}^* + A_1^*\mathrm{y}_{t-1} + \cdots + A_p^*\mathrm{y}_{t-p} + \mathrm{u}_t, \qquad \mathrm{u}_t \sim (0, \Lambda),$$

where Λ is diagonal.

Interpretation

- Restrictions are imposed on A.
- The diagonal matrix Λ carries the structural shock variances.
- Recursive systems are a special case of the A-model.

B-model

In the B-model, the reduced form is written as

$$\mathrm{y}_t = \mathrm{c} + \Phi_1 \mathrm{y}_{t-1} + \cdots + \Phi_p \mathrm{y}_{t-p} + B\mathrm{u}_t, \qquad \mathrm{u}_t \sim (0, I_n),$$

so that

$$\Omega = BB'.$$

Interpretation

- Restrictions are imposed directly on the impact matrix B.
- Lower-triangular B with positive diagonal elements gives a Cholesky-type orthogonalization.

AB-model

The AB-model combines the two:

$$A\mathrm{y}_t = \mathrm{c}^* + A_1^*\mathrm{y}_{t-1} + \cdots + A_p^*\mathrm{y}_{t-p} + B\mathrm{u}_t, \qquad \mathrm{u}_t \sim (0, I_n).$$

Then

$$\varepsilon_t = A^{-1}B\mathrm{u}_t, \qquad \Omega = A^{-1}BB'A^{-1\prime}.$$

Interpretation

Restrictions are split across contemporaneous equations and the impact matrix. This is flexible, but identification becomes more demanding.

Comparison of A-, B-, and AB-models

Model	Main restricted object	Typical use
A-model	A	contemporaneous zero restrictions
B-model	B	impact matrix / Cholesky-type schemes
AB-model	both A and B	richer hybrid identification

What stays common

All three are just different parameterizations of the same goal: recover structural shocks from the reduced-form system.

R note: estimating A- and B-models in vars

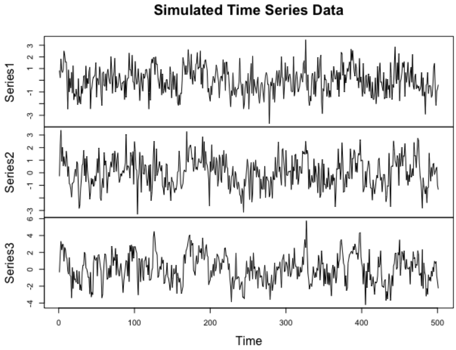
```
library(vars)

var_fit <- VAR(y_mat, p = 1, type = "const")

A_matrix <- matrix(c(1, 0, 0,
                     NA, 1, 0,
                     NA, NA, 1), nrow = 3, byrow = TRUE)

B_matrix <- matrix(c(NA, 0, 0,
                     NA, NA, 0,
                     NA, NA, NA), nrow = 3, byrow = TRUE)

SVAR_A <- SVAR(var_fit, Amat = A_matrix, max.iter = 1000)
SVAR_B <- SVAR(var_fit, Bmat = B_matrix)

irf_A <- irf(SVAR_A, n.ahead = 10, boot = TRUE)
irf_B <- irf(SVAR_B, n.ahead = 10, boot = TRUE)
```

Implementation logic

NA means "estimate this entry"; fixed zeros and ones encode the identifying restrictions.

A simulated SVAR data example

The textbook also includes a simulated trivariate example before fitting A- and B-models.



Pedagogical use

Simulation is valuable because we know the data-generating structure in advance and can see how well the identifying restrictions recover it.

Structural IRFs versus reduced-form IRFs

Once the structural shocks are identified, the impulse responses answer a different question:

$$\frac{\partial y_{t+h}}{\partial u_{jt}}, \qquad h = 0, 1, 2, \ldots$$

This is a response to a **structural shock** u_{jt}, not just to a reduced-form innovation.

- Reduced-form IRFs are useful descriptively.
- Structural IRFs aim at economic interpretation and policy analysis.

Natural stopping point

Once we have identified the mean-dynamics side of the VAR, the next question is whether the shocks themselves have constant variance over time.

Why Chapter 4 begins naturally here

So far, our time-series models have focused on the **conditional mean**:

$$\mathrm{E}(y_t \mid \mathcal{F}_{t-1}).$$

But in finance, risk is often about the **conditional variance**:

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

Empirical stylized fact

Returns often show little linear predictability in the mean, but very strong predictability in volatility.

Typical observation

Large movements tend to be followed by large movements; calm periods tend to be followed by calm periods. This is **volatility clustering**.

A picture of clustered volatility

Even when the mean looks close to zero, the amplitude of movements changes dramatically over time.

Conditional mean dependence versus conditional variance dependence

These two ideas are different:

Mean dynamics

ARMA and VAR models describe how

$$\mathrm{E}(y_t \mid \mathcal{F}_{t-1})$$

depends on the past.

Variance dynamics

ARCH and GARCH models describe how

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

depends on the past.

No contradiction

A process can have almost no linear predictability in the mean and yet have strong predictability in its variance.

i.i.d., martingale difference, and white noise

The chapter carefully distinguishes:

- **i.i.d.**: independent and identically distributed;
- **martingale difference sequence (MDS)**:

$$\mathrm{E}(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0;$$

- **white noise**: zero mean, constant unconditional variance, and no serial correlation.

Important subtlety

White noise does *not* require the conditional variance to be constant.

Consequently

It is possible that

$$\mathrm{Corr}(\varepsilon_t, \varepsilon_{t-k}) = 0$$

while

$$\mathrm{Cov}(\varepsilon_t^2, \varepsilon_{t-k}^2) \neq 0.$$

That is exactly the phenomenon ARCH/GARCH models are built to capture.

Why squared residuals matter

If the mean has already been modelled well, the remaining serial pattern often appears not in ε_t itself but in ε_t^2.

Volatility clustering in one sentence

ε_t may be uncorrelated, but ε_t^2 is serially dependent.

- This is why diagnostics for ARCH effects look at squares.
- It is also why volatility models are often estimated on residuals from a mean model.

Engle's ARCH idea

The key idea is that today's conditional variance depends on yesterday's squared shock.

A convenient representation is

$$\varepsilon_t = \sigma_t \nu_t, \qquad \nu_t \sim \text{i.i.d. } (0, 1),$$

with

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.$$

Interpretation

A large shock yesterday raises the conditional variance today.

ARCH(1) and the AR form for squared shocks

The textbook also writes

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + u_t,$$

where u_t is a new white-noise process.

Meaning

ARCH(1) is an AR(1) model for the *squared* disturbances.

Why this is intuitive

If large squared shocks tend to be followed by large squared shocks, then the conditional variance is persistent even when the shocks themselves have mean zero.

Conditional variance in ARCH(1)

By definition,

$$\sigma_t^2 \equiv \mathrm{E}(\varepsilon_t^2 \mid \mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.$$

To guarantee nonnegativity for all histories, we impose

$$\alpha_0 \geq 0, \qquad \alpha_1 \geq 0.$$

Interpretation of the parameters

- α_0 is the baseline level of conditional variance;
- α_1 measures how strongly yesterday's squared shock feeds into today's volatility.

Unconditional variance of ARCH(1)

If the process is weakly stationary with variance $\sigma^2 = \mathrm{E}(\varepsilon_t^2)$, then

$$\sigma^2 = \alpha_0 + \alpha_1 \mathrm{E}(\varepsilon_{t-1}^2) = \alpha_0 + \alpha_1 \sigma^2.$$

Hence

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1},$$

provided

$$0 \leq \alpha_1 < 1.$$

Interpretation

A stationary ARCH(1) process still has *time-varying* conditional variance, but its *unconditional* variance is constant over time.

ARCH(p)

The ARCH(1) model generalizes to

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2.$$

Equivalently,

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + u_t.$$

Interpretation

An ARCH(p) model remembers volatility shocks for exactly p lags.

Parameter restrictions

To keep conditional variance nonnegative, all α_j must be nonnegative.

Stationarity of ARCH(p)

For weak stationarity, the sum of ARCH coefficients must satisfy

$$\alpha_1 + \alpha_2 + \cdots + \alpha_p < 1.$$

Then

$$\mathrm{Var}(\varepsilon_t) = \frac{\alpha_0}{1 - (\alpha_1 + \cdots + \alpha_p)}.$$

Economic meaning

The total persistence of volatility cannot exceed one if we want a finite unconditional variance.

Why ARCH captures volatility clustering

Suppose a large surprise hits today, so $|\varepsilon_t|$ is large. Then ε_t^2 is large, so the next period's conditional variance satisfies

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \varepsilon_t^2$$

in ARCH(1), or its higher-order analogue in ARCH(p).

Consequence

The model predicts a turbulent period after a turbulent period.

But

Pure ARCH often needs large orders in practice. That motivates GARCH.

Asymptotic weak stationarity of ARCH(1)

Starting from an arbitrary initial value,

$$\mathrm{E}(\sigma_t^2) = \alpha_0 + \alpha_1 \mathrm{E}(\sigma_{t-1}^2).$$

Iterating gives

$$\mathrm{E}(\sigma_t^2) = \alpha_0 \sum_{j=0}^{t-2} \alpha_1^j + \alpha_1^{t-1} \mathrm{E}(\sigma_1^2) \longrightarrow \frac{\alpha_0}{1-\alpha_1}$$

when $|\alpha_1| < 1$.

Meaning

Regardless of the starting value, the expected conditional variance converges to the stationary mean of the variance process.

What ARCH misses

ARCH is conceptually elegant, but there is a practical problem:

- financial volatility is often highly persistent;
- matching that persistence with ARCH alone may require many lags;
- too many lags means too many parameters and too little parsimony.

Natural solution

Let today's volatility depend not only on past squared shocks, but also on its own past values. That is the GARCH idea.

Bollerslev's GARCH generalization

A GARCH(p, q) model is

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2.$$

Difference from ARCH

- ARCH uses only lagged squared shocks.
- GARCH uses lagged squared shocks *and* lagged conditional variances.

The workhorse model: GARCH(1,1)

The most commonly used specification is

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

- α_0 is the long-run baseline variance component,
- α_1 is the impact of new information, or the “ARCH effect”,
- β_1 is the persistence in conditional variance, or the “GARCH effect”.

Reason for popularity

With only three parameters, GARCH(1,1) often captures volatility clustering remarkably well.

Positivity restrictions in GARCH

To keep the conditional variance nonnegative, we usually impose

$$\alpha_0 > 0, \qquad \alpha_j \geq 0, \qquad \beta_k \geq 0.$$

Interpretation

The model builds variance from nonnegative pieces:

- baseline variance,
- past squared shocks,
- and past variances.

Practical note

These are sufficient conditions for positivity in the basic GARCH model and are standard in empirical work.

Unconditional variance of GARCH(1,1)

If ε_t is weakly stationary, then

$$\sigma^2 = \mathrm{E}(\sigma_t^2) = \alpha_0 + \alpha_1\sigma^2 + \beta_1\sigma^2.$$

So

$$\sigma^2 = \frac{\alpha_0}{1-(\alpha_1+\beta_1)},$$

provided

$$\alpha_1 + \beta_1 < 1.$$

Central persistence measure

The quantity $\alpha_1 + \beta_1$ is the fundamental measure of volatility persistence in GARCH(1,1).

GARCH as an ARCH(∞) model

Iterating

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

gives

$$\sigma_t^2 = \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{j=1}^{\infty} \beta_1^{j-1} \varepsilon_{t-j}^2.$$

Interpretation

GARCH(1,1) is an infinite-order ARCH model with geometrically decaying coefficients.

Econometric advantage

This is why GARCH is far more parsimonious than fitting a very high-order ARCH.

How α_1 and β_1 change the volatility dynamics

- Large α_1 means volatility reacts sharply to new shocks.
- Large β_1 means volatility decays slowly once it has risen.
- Large $\alpha_1 + \beta_1$ means volatility is highly persistent.

Heuristic language

- $\alpha_1 =$ impact or news sensitivity,
- $\beta_1 =$ memory,
- $\alpha_1 + \beta_1 =$ persistence.

Weak stationarity of GARCH(p, q)

For the general model

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2,$$

a sufficient and standard condition for weak stationarity is

$$\sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k < 1.$$

Then

$$\mathrm{Var}(\varepsilon_t) = \frac{\alpha_0}{1 - \sum_{j=1}^{p} \alpha_j - \sum_{k=1}^{q} \beta_k}.$$

Interpretation

The total persistence in conditional variance must stay below one for the unconditional second moment to exist.

Persistence and near-integration

In many empirical financial series, estimates satisfy

$$\alpha_1 + \beta_1 \approx 1.$$

Meaning

Volatility shocks decay very slowly:

- a crisis can affect estimated volatility for a long time;
- the variance process is highly persistent;
- the boundary case suggests IGARCH-type behavior.

Transition to next lecture

Lecture 10 returns to the boundary and asymmetric cases, including IGARCH and leverage-type models.

Strict stationarity of GARCH(1,1)

Weak stationarity is about finite second moments. Strict stationarity is a stronger concept.

For

$$\varepsilon_t = \sigma_t \nu_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

Nelson's condition for strict stationarity is

$$\mathrm{E}\left[\log(\beta_1 + \alpha_1 \nu_t^2)\right] < 0.$$

Important point

This condition can hold even when

$$\alpha_1 + \beta_1 \geq 1.$$

So a process can be strictly stationary without being weakly stationary.

Why strict and weak stationarity differ here

Weak stationarity requires the second moment to exist. Strict stationarity does not.

In GARCH

- strict stationarity depends on a log-moment condition;
- weak stationarity depends on the existence of a finite unconditional variance.

Analogy

This is another example, like the Cauchy case discussed earlier in the course, where strict stationarity does not automatically imply weak stationarity.

Mean and variance equations work together

In applications we often combine a mean model and a variance model:

$$y_t = \mu_t + \varepsilon_t, \qquad \varepsilon_t = \sigma_t \nu_t,$$

with, for example,

μ_t from an ARMA or VAR equation, $\qquad \sigma_t^2$ from an ARCH/GARCH equation.

Interpretation

- the mean model explains expected movement;
- the variance model explains risk or uncertainty around that movement.

A short R note: simulating and selecting a GARCH model

```
library(rugarch)

spec <- ugarchspec(
  variance.model = list(model = "sGARCH",
                        garchOrder = c(1, 1)),
  mean.model = list(armaOrder = c(0, 0),
                    include.mean = FALSE)
)

fit <- ugarchfit(spec = spec, data = y)

infocriteria(fit)
sigma_hat <- sigma(fit)
resid_std <- residuals(fit, standardize = TRUE)
```

Interpretation

Estimation details are postponed to Lecture 10, but the basic workflow is already clear: specify the mean, specify the variance, estimate, then inspect persistence and residual diagnostics.

How students should read a fitted GARCH(1,1)

Given estimates $\widehat{\alpha}_0$, $\widehat{\alpha}_1$, $\widehat{\beta}_1$, ask:

1. Is $\widehat{\alpha}_0 > 0$?
2. Are $\widehat{\alpha}_1$ and $\widehat{\beta}_1$ nonnegative?
3. Is $\widehat{\alpha}_1 + \widehat{\beta}_1 < 1$?
4. If the sum is near one, how persistent is volatility?
5. Does the standardized residual still show serial dependence or ARCH effects?

Interpretive summary

A large $\widehat{\beta}_1$ with a moderate $\widehat{\alpha}_1$ is the classic signature of long-lived volatility persistence.

What Lecture 9 has established

- Structural VARs distinguish reduced-form innovations from economically meaningful shocks.
- Identification requires restrictions: short-run, long-run, or parameterized A/B/AB schemes.
- Volatility clustering is dependence in second moments, not necessarily in first moments.
- ARCH models let the conditional variance depend on past squared shocks.
- GARCH models add lagged conditional variance and therefore deliver much richer persistence with few parameters.
- Weak stationarity depends on coefficient sums; strict stationarity in GARCH depends on a log-moment condition.

Preview of Lecture 10

Lecture 10 stays in Chapter 4 and pushes the volatility block further:

- asymmetric volatility models,
- IGARCH and boundary persistence,
- likelihood / quasi-likelihood estimation,
- ARCH-LM and other diagnostics,
- and a fuller R workflow for fitting and checking GARCH models.

Course logic

Lecture 9 gives the theoretical architecture. Lecture 10 turns that architecture into a full empirical toolkit.