

ARMIA and GARCH

Espen Sirnes

2025-04-24

Table of contents

1	Overview	1
2	The ARIMA process	2
3	The GARCH process	4
3.1	Confusing terminology	6
4	Literature	6

1 Overview

The raw, non-stationary residuals \mathbf{u} are defined as:

$$\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}, \quad (1)$$

where \mathbf{X} is an $T \times K$ array of K covariates across N groups and T time periods, $\boldsymbol{\beta}$ is a $K \times 1$ coefficient vector, and \mathbf{y} is the $T \times 1$ dependent variable.

The reasons for handling this problems are

- Serial correlation in \mathbf{u} can understate standard errors, artificially inflating t -statistics.
- Conditional heteroscedasticity violates the constant-variance assumption.
- Converting \mathbf{u} into *i.i.d.* innovations permits valid inference.

The objective is therefore to transform \mathbf{u} — typically serially correlated and heteroscedastic — into uncorrelated, homoscedastic residuals ε/σ by applying

1. **ARIMA filtering** removes mean persistence and unit-roots, delivering the quasi-innovations ε .
2. **GARCH standardisation** scales out time-varying conditional variance, yielding the final innovations $\tilde{\varepsilon}_t = \varepsilon_t/\sigma_t$.

Formally, the ARIMA step (Box et al. 2015) is

$$\varepsilon = \mathbf{MA}^{-1} \cdot \mathbf{AR} \cdot \Delta_d \mathbf{u} \quad (2)$$

where \mathbf{u} is the $T \times 1$ vector, differenced d times by Δ_d . \mathbf{MA} and \mathbf{AR} are $T \times T$ moving-average and autoregressive matrices.

Heteroscedasticity is subsequently corrected by the GARCH (Bollerslev 1986; Engle 1982) normalization:

$$\sigma^2 = \mathbf{GA}_\sigma^{-1} \mathbf{G} + \mathbf{GA}_\sigma^{-1} \mathbf{AR}_\sigma \varepsilon^2 \quad (3)$$

where ε^2 and σ^2 are $T \times 1$ vectors. \mathbf{GA}_σ and \mathbf{AR}_σ embed the GARCH and ARCH lag polynomials.

The matrix \mathbf{G} is a column vector $T \times 1$ with zeros, except the first term is the initial variance.

Let us now look a bit closer on how these transformations are obtained.

2 The ARIMA process

As noted, \mathbf{u} is the $T \times 1$ matrix of residuals. It is linked to the stationary residuals ε via an ARIMA process:

$$\Delta_d \mathbf{u} = \sum_{j=1}^p \rho_j \mathbf{L}^j \Delta_d \mathbf{u} + \varepsilon + \sum_{j=1}^q \lambda_j \mathbf{L}^j \varepsilon \quad (4)$$

where \mathbf{L} is the lag operator matrix, defined as:

$$\mathbf{L} = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ 0 & 1 & 0 & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}_{T \times T} \quad (5)$$

You can verify yourself that premultiplying a column vector by \mathbf{L} shifts its elements down by one period, inserting a zero in the first position. In this way, \mathbf{L} effectively applies a one-period lag to the variable it is multiplied with.

Although Equation 4 may appear complex at first glance—suggesting that “everything depends on everything”—it simplifies neatly when expressed using lag polynomials:

$$\left(\mathbf{I} - \sum_{j=1}^p \rho_j \mathbf{L}^j \right) \Delta_d \mathbf{u} = \left(\mathbf{I} + \sum_{j=1}^q \lambda_j \mathbf{L}^j \right) \boldsymbol{\varepsilon} \quad (6)$$

By compacting the notation and defining the lag polynomials as matrices, the expression becomes even clearer and more concise. The moving-average (MA) component can then be written as:

$$\mathbf{MA} = \mathbf{I} + \sum_{j=1}^q \lambda_j \mathbf{L}^j, \quad (7)$$

and the corresponding autoregressive (AR) term is

$$\mathbf{AR} = \mathbf{I} - \sum_{j=1}^p \rho_j \mathbf{L}^j. \quad (8)$$

The MA matrix typically look like this

$$\mathbf{MA} = \begin{bmatrix} 1 & & & & \\ \lambda_1 & 1 & & & \\ \lambda_2 & \lambda_1 & 1 & & \\ \vdots & \vdots & \ddots & \ddots & \\ \lambda_q & \cdots & \cdots & \lambda_1 & 1 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \quad (9)$$

And the AR matrix typically like this

$$\mathbf{AR} = \begin{bmatrix} 1 & & & & & \\ -\rho_1 & 1 & & & & \\ -\rho_2 & -\rho_1 & 1 & & & \\ \vdots & \vdots & \ddots & \ddots & & \\ -\rho_p & \cdots & \cdots & -\rho_1 & 1 & \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \quad (10)$$

Hence, we can write (Equation 6) in matrix notation as:

$$\mathbf{AR} \cdot \Delta_d \mathbf{u} = \mathbf{MA} \cdot \varepsilon \quad (11)$$

Our goal is to isolate the independent random noise embedded in the differenced residuals $\Delta_d \mathbf{u}$. Recognising that this is a simple linear algebra problem, we can solve for the white-noise sequence ε by premultiplying both sides by the inverse of the moving-average operator \mathbf{MA}^{-1} :

$$\varepsilon = \mathbf{MA}^{-1} \cdot \mathbf{AR} \cdot \Delta_d \mathbf{u} \quad (12)$$

Hence, the white noise error terms ε can be obtained through standard linear algebra by inverting the moving average operator. However, in practice, explicitly computing the inverse is often computationally expensive, so a recursive filtering approach is typically preferred for efficiency.

3 The GARCH process

The primary objective of the GARCH transformation is to correct for serially conditioned heteroskedasticity — that is, time-varying volatility in the error process. Conceptually, the procedure closely parallels the ARIMA approach, but with an inverted goal: whereas ARIMA seeks to eliminate temporal dependence in the residuals by extracting the *innovations*, GARCH focuses on modelling the *volatility* itself.

In this framework, we estimate the serially correlated conditional standard deviation vector σ , which is then used to standardise the stationary ARIMA residuals ε , producing homoscedastic white noise via:

$$\tilde{\varepsilon}_t = \varepsilon_t / \sigma_t \quad (13)$$

This is conceptually the inverse of the ARIMA transformation, which seeks to *remove* serial correlation from the residuals. GARCH, by contrast, models that correlation in volatility.

We begin with a general GARCH formulation. GARCH simply means that the current variance depends on the past variance and the previous squared residuals. This can be written as

$$\sigma^2 = \mathbf{G} + \sum_{j=1}^m \psi_j \mathbf{L}^j \varepsilon^2 + \sum_{j=1}^k \gamma_j \mathbf{L}^j \sigma^2 \quad (14)$$

The matrix \mathbf{G} is a column vector $T \times 1$ with zeros, except the first term is the initial variance. As with ARIMA, we can rearrange the GARCH equation into an explicit solvable form:

$$\left(\mathbf{I} - \sum_{j=1}^k \gamma_j \mathbf{L}^j \right) \sigma^2 = \mathbf{G} + \sum_{j=1}^m \psi_j \mathbf{L}^j h(\varepsilon, c) \quad (15)$$

As with the ARIMA model, the problem becomes more transparent when we express the lag polynomials as matrix operators:

$$\mathbf{GA}_\sigma = \mathbf{I} - \sum_{j=1}^k \gamma_j \mathbf{L}^j \quad (16)$$

$$\mathbf{AR}_\sigma = \sum_{j=1}^m \psi_j \mathbf{L}^j \quad (17)$$

\mathbf{GA}_σ and \mathbf{AR}_σ have similar structures as Equation 9 and Equation 10.

Then, analogous to the ARIMA formulation, we can express the Equation 15 in operator form as:

$$\mathbf{GA}_\sigma \cdot \sigma^2 = \mathbf{G} + \mathbf{AR}_\sigma \cdot \varepsilon^2 \quad (18)$$

As is standard with linear systems, we solve for σ^2 by premultiplying both sides by the inverse of \mathbf{GA}_σ :

$$\sigma^2 = \mathbf{GA}_\sigma^{-1} (\mathbf{G} + \mathbf{AR}_\sigma \cdot \varepsilon^2) \quad (19)$$

This transformation deflates the residuals by their conditional standard deviation, yielding a series ε_t/σ_t that is homoscedastic under the model.

3.1 Confusing terminology

Following convention, the effects from ε^2 are referred to as **autoregressive**, even though they represent the response to past innovations. Historically, the term GARCH was introduced by (Bollerslev 1986) to generalise the ARCH process by adding a persistent variance component.

This naming choice may appear counterintuitive when compared to ARIMA models, where:

- **Autoregressive (AR)** components capture persistence over time.
- **Moving average (MA)** components capture transient innovation shocks.

In GARCH models, however, the roles are reversed:

- **ARCH terms** (lagged squared residuals) reflect immediate volatility responses — akin to MA terms in ARIMA.
- **GARCH terms** (lagged variances) model long-run volatility persistence — functionally closer to AR components.

Thus, it might have been more consistent to label the ARCH component as the moving average part. One possible justification for the current convention is that the ARCH effect is the component inverted when solving the reduced-form variance equation. However, it is crucial to recognise that the dependent variable differs: ARIMA models solve for the signal y (via ε), while GARCH models solve directly for the conditional variance σ^2 .

4 Literature

Bollerslev, Tim. 1986. “Generalized Autoregressive Conditional Heteroskedasticity.” *Journal of Econometrics* 31 (3): 307–27.

- Box, George EP, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica: Journal of the Econometric Society*, 987–1007.