ST418 Week 8

Kaixin Liu[1]
[1]PhD Student in Statistics, LSE

Mar 12, 2026

Table of Contents

Table of Contents

ARIMA model: handling nonstationarity by differencing

Integrated process

A series $\{x_t\}$ is called integrated of order d, written $x_t \sim I(d)$, if

$$\Delta^d x_t = y_t,$$

where $\{y_t\}$ is stationary.

Definition (ARIMA(p, d, q))

If the differenced series $\Delta^d x_t$ follows an ARMA(p, q) model, then

$$x_t \sim \text{ARIMA}(p, d, q).$$

Equivalently,

$$\Phi(B)(1 - B)^d x_t = \alpha + \Theta(B)\varepsilon_t, \qquad \varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2).$$

Seasonality and multiplicative SARIMA

Let $\nabla_s = 1 - B^s$ be the seasonal difference operator with period s.

Definition (Multiplicative seasonal ARIMA, SARIMA)

$\{x_t\}$ follows a seasonal ARIMA model if

$$\Phi_P(B^s)\Phi(B)\nabla_s^D(1-B)^d x_t = \alpha + \Theta_Q(B^s)\Theta(B)\varepsilon_t,$$

where $\varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2)$, which can be shorthanded as

$$x_t \sim \mathrm{ARIMA}(p,d,q) \times (P,D,Q)_s.$$

Examples of (S)ARIMA models

- **ARIMA**$(0, 1, 1)$**:**

  $$(1 - B)x_t = \alpha + (1 + \theta B)\varepsilon_t.$$

  First differencing removes trend/nonstationarity, then MA(1) captures short memory.

- **SARIMA**$(0, 1, 1) \times (0, 1, 1)_{12}$**:**

  $$(1 - B)(1 - B^{12})x_t = \alpha + (1 + \theta B)(1 + \Theta B^{12})\varepsilon_t.$$

  Useful for monthly data with both local dependence and yearly seasonality.

Table of Contents

Yule-Walker estimators for AR(p)

For a weakly stationary AR(p) with mean 0:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma^2),$$

the Yule-Walker equations are

$$\boldsymbol{\Gamma}_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p, \qquad \sigma^2 = s_0 - \boldsymbol{\phi}^\top \boldsymbol{\gamma}_p,$$

where

$$\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top, \qquad \boldsymbol{\gamma}_p = (s_1, \ldots, s_p)^\top.$$

Replacing s_τ by $\widehat{s}_\tau^{(p)}$ gives moment estimators

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p, \qquad \hat{\sigma}^2 = \widehat{s}_0^{(p)} - \hat{\boldsymbol{\phi}}^\top \hat{\boldsymbol{\gamma}}_p.$$

Asymptotic results of Yule-Walker estimators

Theorem (Asymptotic normality of Yule-Walker estimators)

For AR(p),

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2 \mathbf{\Gamma}_p^{-1}), \qquad \sqrt{T}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4).$$

Theorem (Overfitting implication)

If the true order is p but we fit AR(h) with $h > p$, then for redundant high lags (e.g. lag h):

$$\sqrt{T}\,\hat{\phi}_h \xrightarrow{d} N(0, 1).$$

Sample ACF and PACF: practical significance bands

Theorem (Sample ACF/PACF significance bands)

ACF: for MA(q), if $k > q$,

$$\sqrt{T}\,\hat{\rho}_k \xrightarrow{d} N\Big(0,\ 1 + 2\sum_{j=1}^{q} \rho_j^2\Big).$$

95% bands: white-noise case $\hat{\rho}_k \approx \pm 1.96/\sqrt{T}$; more generally

$$\hat{\rho}_k \approx \pm 1.96\sqrt{\frac{1 + 2\sum_{j=1}^{k-1} \hat{\rho}_j^2}{T}}.$$

PACF: under H_0 (AR order $< k$), for $j \geq k$,

$$\sqrt{T}\,\hat{\pi}(j) \xrightarrow{d} N(0, 1),$$

giving an approximate 95% band $\pm 1.96/\sqrt{T}$.

Table of Contents

Motivation: stylized facts of volatility

For returns $\{r_t\}$, volatility is usually modeled as conditional standard deviation:

$$\sigma_t = \sqrt{\mathrm{Var}(r_t \mid \mathcal{F}_{t-1})}.$$

Empirically, many return series show:

1. **Volatility clustering**: high-volatility periods and low-volatility periods cluster.
2. **Continuity**: volatility evolves smoothly over time.
3. **Boundedness**: volatility fluctuates in a stable range. Hence it is usually stationary.
4. **Leverage effect**: volatility response is asymmetric for positive vs negative shocks.

ARCH(p)

Definition (ARCH(p))

The autoregressive conditional heteroscedastic model is

$$x_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_p x_{t-p}^2,$$

where $\varepsilon_t \sim$ i.i.d. $(0,1)$, $\alpha_0 > 0$, and $\alpha_i \geq 0$.

A common condition for covariance stationarity is

$$\sum_{i=1}^{p} \alpha_i < 1.$$

Properties of ARCH

- **Uncorrelated returns**: $\mathbb{E}(x_t) = 0$ and $\mathrm{Cov}(x_t, x_{t-\tau}) = 0$ for $\tau > 0$.
- **Volatility clustering**: large $|x_t|$ tends to increase future conditional variance σ_{t+1}^2.
- **Not independent**: even if x_t is uncorrelated, $\{x_t\}$ is generally not i.i.d.
- **Dependence in squares**: x_t^2 is serially dependent; ACF of squared returns is informative.

For ARCH(1), $\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2$ implies stronger persistence when α_1 is larger.

Limitation of ARCH and need for extension

In practice, volatility persistence is often stronger than what low-order ARCH can capture. To fit long memory in volatility with ARCH alone, one may need a large order p.

This motivates adding lagged conditional variances directly, leading to GARCH.

Table of Contents

GARCH(p, q): motivation and definition

Definition (GARCH(p, q))

$$x_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

where $\varepsilon_t \sim$ i.i.d. $(0, 1)$, $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$.

A standard stationarity condition is

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1.$$

Practical notes on GARCH

- GARCH(1,1) is the workhorse model in empirical finance.
- It captures persistence in volatility better than simple ARCH.
- Squared returns under GARCH admit an ARMA-type representation.
- Standard symmetric GARCH does **not** fully capture leverage effects.

To model asymmetry/leverage, one often moves to EGARCH, GJR-GARCH, etc.