Autoregressive Processes (AR)

Theory, Estimation and Diagnostics

Author

December 1, 2025

Introduction

- Autoregressive (AR) models are fundamental in time series analysis.
- Capture dependence of current value on past values.
- Useful for forecasting and understanding dynamics of stochastic processes.

Definition of AR(p)

- General form:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

- $\epsilon_t \sim WN(0, \sigma^2)$, i.i.d.
- ϵ_t represents the unpredictable component (white noise).

Stationarity Definition

- In general, an AR(p) is stationary if the roots of the characteristic polynomial lie outside the unit circle.

For an autoregressive process AR(p), the characteristic polynomial is defined as:

$$\phi(z) = 1 - a_1 z - a_2 z^2 - \cdots - a_p z^p$$

So, it is stationary if:

$$|z_i| > 1 \quad \forall i$$

This condition ensures that the effect of past shocks decays over time and the process does not diverge.

Stationarity Consequences

As a consequence of stationarity:

1. The **mean** is constant:

$$\mathbb{E}[y_t] = \frac{a_0}{1 - a_1 - \cdots - a_p}$$

2. The **variance** is finite and constant (for AR(1) we have the following, for p>1 solve Yule-Walker system):

$$\mathrm{Var}(y_t) = \frac{\sigma^2}{1 - a_1^2}$$

3. The **autocovariance** depends only on the lag k:

$$\gamma_k = \mathrm{Cov}(y_t, y_{t-k})$$

Supplement: Yule-Walker Equations

The Yule-Walker equations relate the autocovariances of a stationary AR(p) process to its parameters.
For an AR(p) process:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2)$$

The equations are:

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix}$$

- Can be solved to obtain the AR parameters ϕ_i from sample autocovariances.
- Variance of the noise:

$$\sigma^2 = \gamma_0 - \sum_{i=1}^{p} \phi_i \gamma_i$$

Autocovariance and Autocorrelation

- The **autocovariance function** at lag k is defined as:

$$\gamma_k = \mathrm{Cov}(y_t, y_{t-k}) = \mathbb{E}[(y_t - \mu)(y_{t-k} - \mu)]$$

  where $\mu = \mathbb{E}[y_t]$ is the mean of the process.
- Note that $\gamma_0 = \mathrm{Var}(y_t)$ is the variance of the process.
- The **autocorrelation function** (ACF) at lag k is:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

  which measures the correlation between y_t and y_{t-k}.
- For a stationary AR process, γ_k and ρ_k depend only on the lag k, not on t.

Fitting AR(p) OLS: Idea principale

- The OLS method estimates the parameters by **minimizing the sum of squared errors**:

$$S(a_0, a_1, \ldots, a_p) = \sum_{t=p+1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=p+1}^{T} \left(y_t - \sum_{i=0}^{p} a_i y_{t-i} \right)^2$$

- Goal: find parameters $a_0, \ldots, a_p$ that minimize this function.

Fitting AR(p) OLS: Derivation

- Take the derivative of the sum of squared errors w.r.t. each parameter:

$$\frac{\partial S}{\partial a_i} = -2 \sum_{t=p+1}^{T} \left(y_t - \sum_{j=0}^{p} a_j y_{t-j}\right) y_{t-i}$$

- Set each derivative to zero (first-order condition):

$$\frac{\partial S}{\partial a_i} = 0 \quad \forall i = 0, \ldots, p$$

- Solving this system gives the OLS estimates.

Fitting AR(p) OLS: Matrix Form

- Define the response vector Y and design matrix X:

$$Y = \begin{bmatrix} y_p \\ y_{p+1} \\ \vdots \\ y_T \end{bmatrix}, \quad X = \begin{bmatrix} 1 & y_{p-1} & \cdots & y_0 \\ 1 & y_p & \cdots & y_1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_{T-1} & \cdots & y_{T-p} \end{bmatrix}$$

- OLS solution:

$$\hat{a} = (X^\top X)^{-1} X^\top Y$$

- Provides the estimates of all AR parameters at once.

Maximum Likelihood Estimation (MLE)

- The MLE finds parameter values that maximize the probability of observing the given data.

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta \mid \text{data})$$

or equivalently using the log-likelihood:

$$\hat{\theta} = \arg\max_{\theta} \ell(\theta \mid \text{data})$$

- MLE provides consistent and asymptotically efficient estimates under regularity conditions.

AR(p) with Gaussian Innovations

For an AR(p) process:

$$y_t = a_0 + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

The log-likelihood function is:

$$\ell(a_0, \ldots, a_p, \sigma^2) = -\frac{T-p}{2} \log(2\pi) - \frac{T-p}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=p}^{T} (y_t - \mu_t)^2$$

where

$$\mu_t = a_0 + a_1 y_{t-1} + \cdots + a_p y_{t-p}$$

AR(p) with Student-t Innovations

$$y_t = a_0 + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim t_\nu(0, \sigma^2)$$

Log-likelihood:

$$\ell(a_0, \ldots, a_p, \sigma^2, \nu) = \sum_{t=p}^{T} \log \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu\sigma^2}} \left(1 + \frac{(y_t - \mu_t)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}} \right]$$

- MLE requires numerical optimization (e.g., `scipy.optimize.minimize`).
- As $T \to \infty$, MLEs converge to true parameters under regularity conditions.
- Student-t innovations handle heavy-tailed shocks, unlike Gaussian.

Residual Analysis

- Residuals:

$$\hat{\epsilon}_t = y_t - \hat{y}_t$$

- Purpose: assess model fit and detect misspecification
- Check distribution:
  - Histogram: visually inspect normality
  - QQ-plot: compare residual quantiles to theoretical distribution
- Check moments:
  - Mean ≈ 0
  - Variance $\approx \sigma^2$
- Check autocorrelation:
  - ACF of residuals should be near zero for all lags
  - Significant autocorrelation indicates unmodeled structure

Practical Considerations

- Diagnostics are crucial:
  - Residual plots
  - Stationarity checks (unit root tests, ACF decay)
  - Information criteria (AIC, BIC) for model selection
- Limitations:
  - AR models assume linearity and stationarity
  - May not capture heavy tails or volatility clustering
- For heavy-tailed shocks, consider Student-t innovations or GARCH extensions

References

Klaus Neusser, *Time Series Econometrics*.