Notes: Intro to Time Series and Forecasting – Ch5 Modeling and Forecasting with ARMA Processes

Yingbo Li

05/04/2019

Table of Contents

Parameter estimation for ARMA(p, q)

- **When the orders p, q are known**, estimate the parameters

$$\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p), \quad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_q), \quad \sigma^2$$

  - There are $p + q + 1$ parameters in total

- Preliminary estimations
  - Yule-Walker and Burg's algorithm: good for AR(p)
  - Innovation algorithm: good for MA(q)
  - Hannan-Rissanen algorithm: good for ARMA(p, q)

- More efficient estimation: MLE

- **When the orders p, q are unknown**, use model selection methods to select orders
  - Minimize one-step MSE: FPE
  - Penalized likelihood methods: AIC, AICC, BIC

Yule-Walker equations

- $\{X_t\}$ is a casual AR(p) process

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

- Multiplying each side by $X_t, X_{t-1}, \ldots, X_{t-p}$, respectively, and taking expectation, we got the Yule-Walker equations

$$\sigma^2 = \gamma(0) - \phi_1\gamma(1) - \cdots \phi_p\gamma(p)$$

$$\underbrace{\begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{bmatrix}}_{\mathbf{\Gamma}_p} \underbrace{\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}}_{\boldsymbol{\phi}} = \underbrace{\begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{bmatrix}}_{\boldsymbol{\gamma}_p}$$

- Vector representation

$$\mathbf{\Gamma}_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p, \quad \sigma^2 = \gamma(0) - \boldsymbol{\phi}' \boldsymbol{\gamma}_p$$

Yule-Walker estimator and its properties

- Yule-Walker estimators $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \cdots, \hat{\phi}_p)$ are obtained by **solving the hatted version of the Yule-Walker equations**

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p, \quad \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\phi}}' \hat{\boldsymbol{\gamma}}_p$$

- The fitted model is causal and $\hat{\sigma}^2 \geq 0$

$$X_t = \hat{\phi}_1 X_{t-1} + \cdots + \hat{\phi}_p X_{t-p} + Z_t, \quad Z_t \sim \mathrm{WN}(0, \hat{\sigma}^2)$$

- Asymptotic normality

$$\hat{\boldsymbol{\phi}} \dot{\sim} \mathrm{N}\left(\boldsymbol{\phi}, \frac{\sigma^2 \boldsymbol{\Gamma}_p^{-1}}{n}\right)$$

Yule-Walker estimator is a moment estimator: because it is obtained by equating theoretical and sample moments

- Usually moment estimators have much higher variance than MLE
- But Yule-Walker estimators of AR(p) process have the same asymptotic distribution as the MLE
- Moment estimators can fail for MA(q) and general ARMA
  - For example, MA(1): $X_t = Z_t + \theta Z_{t+1}$ with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$.

$$\gamma(0) = (1+\theta^2)\sigma^2, \quad \gamma(1) = \theta\sigma^2 \quad \Longrightarrow \quad \rho(1) = \frac{\theta}{1+\theta^2}$$

Moment estimator of θ is obtained by solving

$$\hat{\rho}(1) = \frac{\hat{\theta}}{1+\hat{\theta}^2} \quad \Longrightarrow \quad \hat{\theta} = \frac{1 \pm \sqrt{1-4\hat{\rho}(1)^2}}{2\hat{\rho}(1)}$$

This can yield complex $\hat{\theta}$ if $|\hat{\rho}(1)| > 1/2$, which can happen if $\rho(1) = 1/2$, i.e., $\theta = 1$

Innovations algorithm: estimate MA coefficients

- Fitted innovations MA(m) model

$$X_t = Z_t + \hat{\theta}_{m1} Z_{t-1} + \cdots + \cdots + \hat{\theta}_{mm} Z_{t-m}, \quad \{Z_t\} \sim \mathrm{WN}(0, \hat{v}_m)$$

  where $\hat{\boldsymbol{\theta}}_m$ and $\hat{v}_m$ are from the innovations algorithm with ACVF replaced by the sample ACVF

- For a MA(q) process, the innovations algorithm estimator $\hat{\boldsymbol{\theta}}_q = (\hat{\theta}_{q1}, \ldots, \hat{\theta}_{qq})'$ is NOT consistent for $(\theta_1, \ldots, \theta_q)'$
- Choice of m: increase m until the vector $(\hat{\theta}_{m1}, \ldots, \hat{\theta}_{mq})'$ stabilizes

Likelihood function of a Gaussian time series

- Suppose $\{X_t\}$ is a Gaussian time series with mean zero
- Assume that covariance matrix $\mathbf{\Gamma}_n = E(\mathbf{X}_n \mathbf{X}_n')$ is nonsingular
- One-step predictors using innovations algorithm: $\hat{X}_1 = 0$ and

$$\hat{X}_{j+1} = P_j X_{j+1}$$

  with MSE $v_j = E\left(X_{j+1} - \hat{X}_{j+1}\right)^2$

  - Example: AR(1)

$$\hat{X}_j = \begin{cases} 0, & j = 1 \\ \phi \hat{X}_{j-1} & j \geq 2 \end{cases}, \quad v_j = \begin{cases} \frac{\sigma^2}{1-\phi^2}, & j = 0 \\ \sigma^2 & j \geq 1 \end{cases}$$

- Likelihood function

$$\begin{aligned} L &\propto |\mathbf{\Gamma}_n|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{X}_n' \mathbf{\Gamma}_n^{-1} \mathbf{X}_n\right) \\ &= (v_0 v_1 \cdots v_{n-1})^{-1/2} \exp\left[-\frac{1}{2} \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{v_{j-1}}\right] \end{aligned}$$

Maximum likelihood estimation of ARMA(p, q)

- Innovations MSE $v_j = \sigma^2 r_j$, where r_j depends on ϕ and $\boldsymbol{\theta}$
- Maximizing the likelihood is equivalent to minimizing

$$-2 \log L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = n \log(\sigma^2) + \sum_{j=1}^{n} \log(r_{j-1}) + \frac{S(\boldsymbol{\phi}, \boldsymbol{\theta})}{\sigma^2},$$

where

$$S(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$$

- MLE $\hat{\sigma}^2$ can be expressed with MLE $\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}$

$$\hat{\sigma}^2 = \frac{S\left(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}\right)}{n}$$

- MLE $\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}$ are obtained by minimizing

$$\log\left[\frac{S(\boldsymbol{\phi}, \boldsymbol{\theta})}{n}\right] + \frac{1}{n} \sum_{j=1}^{n} \log(r_{j-1})$$

Not depend on σ^2!

Asymptotic normality of MLE

- When n is large, for a causal and invertible ARMA(p, q) process,

$$\begin{bmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\boldsymbol{\theta}} \end{bmatrix} \dot{\sim} \mathrm{N}_{p+1}\left(\begin{bmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\boldsymbol{\theta}} \end{bmatrix}, \frac{\mathbf{V}}{n} \right)$$

- For an AR(p) process, MLE has the same asymptotic distribution as the Yule-Walker estimator

$$\mathbf{V} = \sigma^2 \boldsymbol{\Gamma}_p^{-1} \quad \Longrightarrow \quad \hat{\boldsymbol{\phi}} \dot{\sim} \mathrm{N}\left(\boldsymbol{\phi}, \frac{\sigma^2 \boldsymbol{\Gamma}_p^{-1}}{n} \right)$$

Examples of $\mathbf{V}$

- AR(1)

$$\mathbf{V} = 1 - \phi_1^2$$

- AR(2)

$$\mathbf{V} = \begin{bmatrix} 1 - \phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1 - \phi_2^2 \end{bmatrix}$$

- MA(1)

$$\mathbf{V} = 1 - \theta_1^2$$

- MA(2)

$$\mathbf{V} = \begin{bmatrix} 1 - \theta_2^2 & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & 1 - \theta_2^2 \end{bmatrix}$$

- ARMA$(1,1)$

$$\mathbf{V} = \frac{1+\phi\theta}{(\phi+\theta)^2} \begin{bmatrix} (1-\phi^2)(1+\phi\theta) & -(1-\theta^2)(1-\phi^2) \\ -(1-\theta^2)(1-\phi^2) & (1-\phi^2)(1+\phi\theta) \end{bmatrix}$$

Order selection

- Why? Harm of using too large p, q to fit models:
  - Large errors arising from parameter estimation of the model
  - Large MSEs of forecasts
- FPE: only for AR(p) processes

$$\text{FPE} = \hat{\sigma}^2 \frac{n+p}{n-p}$$

- AIC: for ARMA(p, q); approximate Kullback-Leibler discrepancy of the fitted model and the true model, a penalized likelihood method

$$\text{AIC} = -2\log(\hat{L}) + 2(p+q+1)$$

- AICC: for ARMA(p, q); a bias-corrected version of AIC, a penalized likelihood method

$$\text{AICC} = -2\log(\hat{L}) + 2(p+q+1) \cdot \frac{n}{n-p-q-2}$$

Residuals and rescaled residuals

- Residuals of an ARMA(p, q) process

$$\hat{W}_t = \frac{X_t - \hat{X}_t\left(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}\right)}{\sqrt{r_{t-1}\left(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}\right)}}, \quad t = 1, \ldots, n$$

  - Residuals $\{\hat{W}_t\}$ should be similar to white noises $\{Z_t\}$

- Rescaled residuals

$$\hat{R}_t = \frac{\hat{W}_t}{\hat{\sigma}}, \quad \hat{\sigma} = \sqrt{\frac{\sum_{t=1}^n \hat{W}_t^2}{n}}$$

  - Residuals residuals should be approximately $\mathrm{WN}(0, 1)$

Residual diagnostics

1. Plot $\{\hat{R}_t\}$ and look for patterns
2. Compute the sample ACF of $\{\hat{R}_t\}$
   - It should be close to the $\mathrm{WN}(0, 1)$ sample ACF
3. Apply Chapter 1 tests for IID noises

References

- Brockwell, Peter J. and Davis, Richard A. (2016), *Introduction to Time Series and Forecasting, Third Edition*. New York: Springer