Estimação no Domínio do tempo: Covariâncias e modelos ARIMA

Airlane Pereira Alencar

25 de março de 2025

Índice

Ajuste de modelos ARIMA(p,d,q).
Análise descritiva: série veio de processo estacionário?

- Tirar tendência e sazonalidade determinística e ”efeitos”de outras var.
  $X_t = \mu_t + e_t$, em que e_t é estacionário talvez ARMA
  $\mu_t = tendencia(t) + sazon(t) + tend2(t)$
  Exemplo: Produção industrial com mudança de tendência em 2008 e 2014 + AR(4)
  Evitar regressão espúria.
- Se parece ter tendência estocástica: Diferenças
  $Z_t = \Delta X_t = X_t - X_{t-1} = (1 - B)X_t$ ou
  $Z_t = \Delta^{12} X_t = (1 - B^{12}) = X_t - X_{t-12}$
  Se tira uma diferença e ΔX_t parece estacionária, dizemos $X_t \sim I(1)$.

Metodologia

- **Especificação**: Definir classe de modelos. ex:ARIMA(p,d,q)
- **Identificação**: Tirar tendências e com base na FAC e FACP propor (p,q)
- **Estimação** dos parâmetros
- **Verificação**: Análise de resíduos padronizados como RB, gaussiano
- Depois, inferência e **previsão**

ARMA(p,q)

$$
\begin{aligned}
\tilde{Z}_t &= \phi_1 \tilde{Z}_{t-1} + \ldots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} \ldots - \theta_q a_{t-q} \\
(1 - \phi_1 B - \cdots - \phi_p B)\tilde{Z}_t &= (1 - \theta_1 B \ldots - \theta_q B^q) a_t \\
\Phi(B)\tilde{Z}_t &= \Theta(B) a_t
\end{aligned}
$$

com $a_t \sim RB$

- Estacionário: se as raízes de $\Phi(B) = 0$ estiverem fora do círculo unitário
- Invertível: se as raízes de $\Theta(B) = 0$ estiverem fora
- FAC e FACP parece cair para zero e deve propor valores (p,q) até ter resíduo como RB
- Deve propor valores (p,q) até ter resíduo como RB

Modelos ARIMA(p,d,q)

X_t não estac. $\Rightarrow Z_t = \Delta^d X_t$ estacionário $\Rightarrow Z_t \sim ARMA(p,q)$
$\tilde{Z}_t = Z_t - \mu$.

$$\begin{aligned}\tilde{Z}_t &= \phi_1 \tilde{Z}_{t-1} + \ldots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots \theta_q a_{t-q} \\ (1 - \phi_1 B - \ldots)\tilde{Z}_t &= (1 - \theta_1 B - \ldots \theta_q B^q) a_t \\ \Phi(B)\tilde{Z}_t &= \Theta(B) a_t,\end{aligned}$$

$a_t \sim RB(0, \sigma^2)$.

- Estimação de ARIMA consiste na estimação do ARMA(p,q)
- Máxima verossimilhança cond.: $a_t \sim N(0, \sigma_a^2)$ indep.
- Generalizações: outras dist como a t, variância não constante, inclusão de covariáveis.

Estimação ARIMA - Verossimilhança Condicional

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \ldots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

$$\begin{aligned} E(\tilde{Z}_t|past) &= \phi_1 \tilde{z}_{t-1} + \ldots + \phi_p \tilde{z}_{t-p} - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \\ Var(\tilde{Z}_t|past) &= \sigma_a^2 \\ \tilde{Z}_t|past &\sim N(E(\tilde{Z}_t|past), \sigma_a^2) \end{aligned}$$

Verossimilhança Condicional a valores iniciais de $\tilde{Z}_t$'s e a's (MT).

$$\begin{aligned} L(\boldsymbol{\theta}) &= f(\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_T | valores\ iniciais) = \prod_{t=1}^{T} f_{\boldsymbol{\theta}}(\tilde{z}_t|\tilde{z}_{t-1}, \ldots) \\ &= \left[\frac{1}{\sqrt{2\pi\sigma_a^2}}\right]^T exp\left\{-\sum_{t=1}^{T} \frac{[\tilde{Z}_t - E(\tilde{Z}_t|past)]^2}{2\sigma_a^2}\right\} \\ &= \left[\frac{1}{\sqrt{2\pi\sigma_a^2}}\right]^T exp\left\{-\sum_{t=1}^{T} \frac{a_t^2}{2\sigma_a^2}\right\} \end{aligned}$$

Estimação ARIMA

Log-Verossimilhança Condicional

$$
\begin{aligned}
l(\boldsymbol{\theta}) &= -\frac{T}{2} ln(\sigma_a^2) - \frac{S(\eta|\tilde{z}_t, past)}{2\sigma_a^2} \\
S(\eta|\tilde{z}_t, past) &= \sum_{t=1}^{T} a_t^2(\eta|\tilde{\mathbf{Z}}, past)
\end{aligned}
$$

$\eta = (\phi, \boldsymbol{\theta})$ em MT (poderia incluir μ ou intercepto e efeitos)

- Escolha dos p valores iniciais $\tilde{Z}_0, \tilde{Z}_{-1}, \ldots, \tilde{Z}_{-p+1}$ e q valores iniciais de $a_t (= E(a_t) = 0)$
- Predizer a_t fixando η (ex.7.1. MT(2018))

Estimação ARIMA

- Escolha dos p valores iniciais $\tilde{Z}_0, \tilde{Z}_{-1}, \ldots, \tilde{Z}_{-p+1}$ e q valores iniciais de $a_t (= E(a_t) = 0)$
  ex: $Y_t = c + \phi_1 Y_{t-1} - \theta_1 a_{t-1} + a_t,\ a_t \sim iidN(0, \sigma^2)$
  Precisa fixar os valores iniciais y_0 e a_0.
  Prediz a sequência de erros usando os estimadores iniciais de $\eta = (c, \phi, \theta)$:
  $a_t = y_t - c - \phi y_{t-1} + \theta a_{t-1}$

$$lnL(\Theta) = -\frac{T}{2} ln(2\pi) - \frac{T}{2} ln(\sigma^2) - \sum_{t=1}^{T} \frac{(y_t - c - \phi y_{t-1} + \theta a_{t-1})^2}{2\sigma^2}$$

  Atualiza os valores dos parâmetros, recalcula os erros,....
- Predizer a_t fixando parâmetro η (ex.7.1. MT(2018))

Estimação ARIMA

- Escolha dos p valores iniciais $\tilde{Z}_0, \tilde{Z}_{-1}, \ldots, \tilde{Z}_{-p+1}$ e q valores iniciais de $a_t (= E(a_t) = 0)$
  ex: $Y_t = c + \beta x_t + e_t$, $e_t = \phi_1 e_{t-1} - \theta_1 a_{t-1} + a_t,\ a_t \sim iidN(0, \sigma^2)$
  Precisa fixar os valores iniciais y_0 e a_0.
  Prediz a sequência de erros usando os estimadores iniciais de $\eta = (c, \beta, \phi, \theta)$:
  $e_t = y_t - c - \beta x_t$

$$lnL(\Theta) = -\frac{T}{2} ln(2\pi) - \frac{T}{2} ln(\sigma^2) - \sum_{t=1}^{T} \frac{(y_t - c - \beta x_t - \phi y_{t-1} + \theta a_{t-1})^2}{2\sigma^2}$$

  Atualiza os valores dos parâmetros, recalcula os erros,....
- Predizer a_t fixando η (ex.7.1. MT(2018))

Estimação ARIMA

- **Não condicional**
  - Agora, Esperança não condicional ao passado

$$[a_t(\eta|\tilde{\mathbf{Z}}, past)] = E[(a_t|\eta, W)]$$

  - Precisa de backforecasting, usando equivalência entre

$$\mathbf{\Phi}(B)W_t = \mathbf{\Theta}(B)a_t \quad \mathbf{\Phi}(F)W_t = \mathbf{\Theta}(F)e_t$$

- Verossimilhança exata: AR(1) em MT inclui parâmetro para média de Z_0, Z_{-1}

Variância dos estimadores

$\boldsymbol{\eta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$ é $k \times 1$, $k = p + q$
Para n grande, estimador de MV

$$\widehat{\eta} \quad {}^{D}\rightarrow \quad N_k(\eta, \mathbf{V}) \tag{1}$$

$$\mathbf{V} = 2\sigma_a^2 \begin{bmatrix} \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_1^2} & \cdots & \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_1 \partial \eta_k} \\ \vdots & & \vdots \\ \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_k \partial \eta_k} & \cdots & \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_n^2} \end{bmatrix}^{-1} \tag{2}$$

EMV de σ^2 é $\widehat{\sigma}_a^2 = \frac{S(\widehat{\eta})}{n}$
Para n grande $\widehat{\sigma}^2$ e $\widehat{\eta}$ são não correlacionados
Ex: AR(1): $var(\widehat{\phi}) \approx \frac{1-\phi^2}{n}$ e MA(1): $var(\widehat{\theta}) \approx \frac{1-\theta^2}{n}$

Modelo espaço de estados

- Os modelos ARMA podem ser escritos como modelo espaço de estados.
- Filtro de Kalman para predizer as variáveis latentes
- Arima da library(forecast)
- Propor prioris e estimar maximizando a dist. a posteriori
- Estimar usando Arima da library(forecast) com exemplo Airlines.
- Podem tentar só ajustar escrevendo a verossimilhança em planilha e maximizar usando solver.

Referências

All Time series analysis

- Morettin e Toloi
- Shumway and Stoffer
- Wei
- Hyndman. Forecasting