EAE1223: Econometria III

Aula 5 - Metodologia de Box-Jenkins

Luis A. F. Alvarez

28 de março de 2024

A metodologia

- A metodologia de Box-Jenkins consiste numa série de etapas para estimar um modelo univariado de previsão.
    - Ideia é estimar um modelo simples, embora flexível, aos dados.
- Trata-se da metodologia básica de previsão em séries de tempo.
    - Diversas metodologias modernas incorporam o “espírito” de Box-Jenkins.
    - *Benchmark* para avaliar outros modelos de previsão.

Etapas da metodologia

- A metodologia consiste de quatro etapas:
    1. Identificação: nessa etapa, avaliamos os dados e identificamos quais modelos são candidatos plausíveis para reproduzir os dados.
    2. Estimação: nessa etapa, estimamos os modelos candidatos.
    3. Diagnóstico: nessa etapa, avaliamos quais dos modelos se saíram melhor, de acordo com alguns critérios.
    4. Previsão: por fim, realizamos a previsão de acordo com nosso modelo.
- Nesta aula, discutiremos cada uma dessas etapas.
- Começaremos revisando a classe de modelos estudadas na metodologia de Box-Jenkins.

Modelos considerados

MA(q)

- Dizemos que uma série de tempo $\{Y_t\}$ segue um MA(q) se:

$$Y_t = \alpha + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

  onde $\{\epsilon_t\}$ é um ruído branco.

- Série hoje depende diretamente da realização atual e das últimas q realizações de um ruído branco.

- **Todo** processo de média móvel é fracamente estacionário.
  - De fato, $\mathbb{E}[Y_t] = \alpha$, $\mathbb{V}[Y_t] = \left(1 + \sum_{j=1}^{q} \theta_j^2\right) \sigma_\epsilon^2$, $\text{cov}(Y_t, Y_{t-s}) = (\theta_s + \sum_{j=s+1}^{q} \theta_j \theta_{j-s}) \sigma_\epsilon^2$ se $s \leq q$ e 0 do contrário. Correlação morre após q períodos.

- Um processo MA(q) é dito **invertível** se pode ser escrito como:

$$Y_t = \omega + \sum_{j=1}^{\infty} \psi_j Y_{t-j} + \epsilon_t$$

  (MA(q) pode ser representado como AR(∞)).

AR(p) estacionário

- Dizemos que uma série de tempo $\{Y_t\}$ segue um AR(p) estacionário se ela se escreve como:

$$Y_t = \alpha + \sum_{j=1}^{p} \beta_j Y_{t-j} + \epsilon_t$$

onde $\{\epsilon_t\}$ é um ruído branco e os coeficientes β_j são tais que o processo resultante é fracamente estacionário.

- Série hoje depende diretamente das realizações passadas nos últimos p períodos, mais um ruído branco.
  - Mas persistência não é tão grande, de modo que a série é estacionária (não há raiz unitária).

- Recorde-se que um AR(p) é estacionário se, e somente se, ele se escreve como um MA(∞):

$$Y_t = \kappa + \epsilon_t + \sum_{j=1}^{\infty} \tau_j \epsilon_{t-j}$$

ARMA(p,q) estacionário

- Dizemos que uma série de tempo $\{Y_t\}$ segue um ARMA(p,q) estacionário se:

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

  onde $\{\epsilon_t\}$ é um ruído branco, e os β_i são tais que o processo resultante é estacionário.

- Combinação dos dois modelos anteriores.

- Metodologia de Box-Jenkins visará a estimar modelos na classe ARMA(p,q).

E se os dados forem não estacionários?

- Até agora, discutimos a modelagem supondo as séries estacionárias.
  - Como fazer a previsão em casos não estacionários?
- Se as séries apresentarem uma tendência estocástica, trabalhamos com os dados em primeira diferença $\{\Delta y_t\}$.
- Conduzidas as etapas da metodologia Box-Jenkins com os dados diferenciados, e encontradas projeções para Δy_t fora da amostra, recompomos as projeções em nível usando o fato de que $y_{t+1} = y_t + \Delta y_{t+1}$.
  - Isto é, se temos T observações, projetamos $\widehat{y_{T+1}} = y_t + \widehat{\Delta y_{T+1}}$, $y_{t+2} = y_t + \widehat{\Delta y_{t+1}} + \widehat{\Delta y_{t+2}}$, e assim por diante.

Modelos ARIMA(p,d,q)

- Em outras palavras, no caso de raiz unitária, a metodologia irá estimar um modelo ARIMA(p,1,q), da forma:

$$\Delta y_t = \alpha + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

- De modo geral, se a série é I(d), a modelagem considerará modelos ARIMA(p,d,q):

$$\Phi(L)(1-L)^d y_t = \alpha + \Psi(L)\epsilon_t \,,$$

para polinômios $\Phi(L)$ e $\Psi(L)$ de grau p e q, respectivamente.

Previsão com tendência determinística

- Se séries apresentam tendência determinística, conduzimos a metodologia com dados *detrended*. Calculadas as projeções *detrended*, recompomos projeções em nível usando a tendência estimada.

- Por exemplo, se estimamos uma tendência linear:

$$y_t = \tilde{a} + \tilde{b}t + \tilde{\xi}_t$$

  e ajustamos um modelo ARMA para $\tilde{\xi}_t$, a projeção para fora da amostra é:

$$\widehat{y_{T+h}} = \tilde{a} + \tilde{b}(T+h) + \widehat{\tilde{\xi}_{T+h}},$$

  onde $\widehat{\tilde{\xi}_{T+h}}$ é a projeção do ARMA para $T+h$ (veremos como calculá-la na etapa de previsão).

- Estimação de ARMA para dado *detrended* é equivalente a estimar um modelo ARMA(p,q) com tendência determinística.

$$\Phi(L)y_t = \alpha + \gamma t + \Psi(L)\epsilon_t\,.$$

“Filosofia” da metodologia de Box-Jenkins

- A restrição a modelos ARMA(p,q) pode ser entendida a partir do teorema de decomposição de Wold.
- Segundo esse teorema, **qualquer processo fracamente estacionário** pode ser representado pela soma de um MA(∞), acrescido de uma função determinística dos y no passados:

$$y_t = \epsilon_t + \sum_{l=0}^{\infty} \psi_l \epsilon_{t-l} + \kappa_t \,,$$

  onde κ_t é função “aproximadamente” linear de y_{t-1}, $y_{t-2} \ldots$

- Ideia de Box-Jenkins é aproximar essa representação por um ARMA(p,q), com p e q pequenos.
  - Ideia é que aproximação parcimoniosa, por ser menos ruidosa, tende a funcionar melhor que modelos muito complexos.

Identificação

Identificação de um ARMA(p,q)

- A etapa de identificação da metodologia Box-Jenkins consiste em encontrar quais modelos da classe ARMA(p,q) melhor caracterizam a série de interesse.

- A identificação consiste em analisar a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) da série estacionária.

Função de autocorrelação serial (FAC)

- A função de autocorrelação (FAC) de uma série $\{Y_t\}_t$ estacionária é o mapa que associa, a cada número $k > 0$, a autocorrelação de ordem k, i.e. $\gamma_k = \text{cor}(Y_t, Y_{t-k})$.
- Note que essa função está bem definida para processos estacionários, visto que $\text{cor}(Y_t, Y_{t-k})$ não depende de t.

$$\text{cor}(Y_t, Y_{t-k}) = \frac{\text{cov}(Y_t, Y_{t-k})}{\text{sd}(Y_t)\,\text{sd}(Y_{t-k})} = \frac{\text{cov}(Y_t, Y_{t-k})}{\mathbb{V}(Y_t)}$$

Função de autocorrelação serial parcial (FACP)

- A função de autocorrelação parcial (FACP) de uma série $\{Y_t\}_t$ estacionária é o mapa que associa, a cada número $k > 0$, a correlação θ_k entre Y_t e Y_{t-k}, controlando por $Y_{t-1}, Y_{t-2}, \ldots Y_{t-k+1}$.

- A autocorrelação parcial de ordem k (θ_k) é dada pelo coeficente τ_k associado a Y_{t-k} no modelo preditivo linear:

$$\begin{aligned} Y_t = \tau_0 + \tau_1 Y_{t-1} + \tau_2 Y_{t-2} \ldots + \tau_k Y_{t-k} + \nu_t \\ \mathbb{E}[\nu_t] = 0, \mathbb{E}[\nu_t Y_{t-j}] = 0, \quad j = 1, \ldots k \end{aligned} \tag{1}$$

- Note que função está bem definida para processos estacionários, visto que coeficientes do melhor preditor linear em (1) não dependem de t.

FAC E FACP ESTIMADAS

- Na prática, não observamos a FAC nem a FACP de um processo, mas podemos estimá-las usando as realizações da série de interesse.
  - Estimamos a FAC calculando as autocorrelações nos dados.

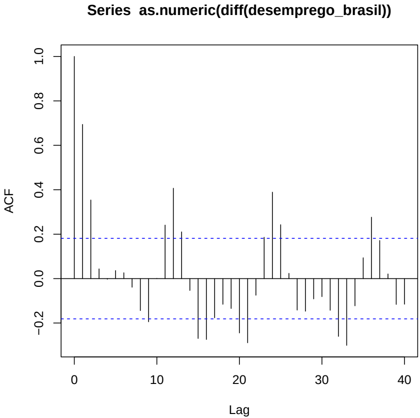
$$\hat{\gamma}_k = \frac{\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2},$$

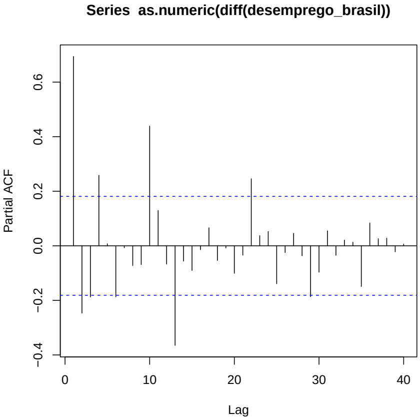
  onde $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$.
  - Estimamos a FACP ajustando o modelo (1) aos dados.

Inferência sobre FAC e FACP populacionais

- Como observamos apenas algumas realizações do processo, gostaríamos de testar hipóteses sobre a FAC e FACP populacionais.
- Sob algumas condições mais estringentes, o intervalo $[-2/\sqrt{T}, 2/\sqrt{T}]$ é uma região de aceitação aproximadamente válida para o teste da nula $\gamma_k = 0$ ($\theta_k = 0$ na FACP) contra a alternativa bilateral ao nível de significância de 5%.
  - São esses intervalos que são apresentados, no R, quando computamos a FAC e FACP.
  - Para autocorrelações (parciais) estimadas que excedem esses limites, rejeitamos a hipótese nula de não autocorrelação (parcial) a essa ordem.

FAC e FACP amostrais de ΔDesemprego





Inferência conjunta sobre a FAC

- Para testar a nula conjunta de que $\gamma_1 = \gamma_2 = \ldots = \gamma_s = 0$, onde s é pequeno relativamente a T, podemos usar a estatística de Ljung-Box:

$$\hat{Q} = T(T+2)\sum_{r=1}^{s} \frac{\hat{\gamma}_r^2}{T-k}\,.$$

- Com T grande, sob a nula, $\hat{Q}$ segue uma qui-quadrado com s graus de liberdade. Valores altos da estatística são evidência contra a nula, i.e. evidência de que ao menos uma autocorrelação entre as testadas é diferente de zero.

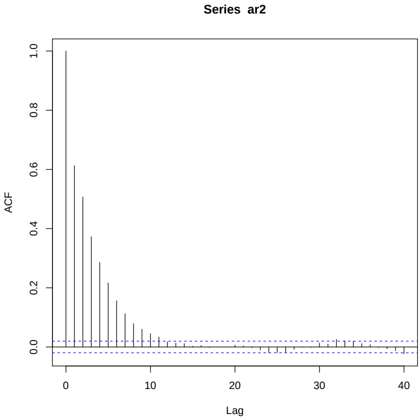
FAC e FACP de um AR(p) estacionário

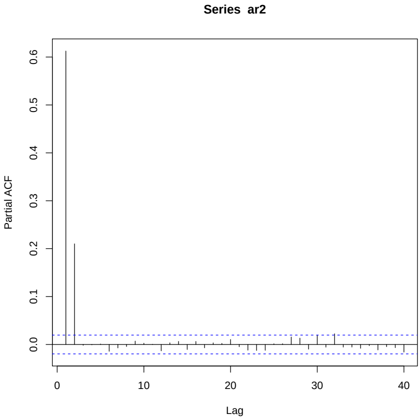
- Como vimos em aula anterior, a FAC de um AR(1) estacionário é dada por:
$$\gamma_k = \text{cor}(Y_t, Y_{t-k}) = \beta_1^k, \quad |\beta_1| < 1 \tag{2}$$
isto é, a FAC apresenta decaimento geométrico em direção a zero.

- De modo geral, a FAC de um AR(p) estacionário apresenta decaimento em direção a zero, visto que um AR(p) estacionário pode ser escrito como um MA(∞), i.e.
$$Y_t = \mu + \epsilon_t + \sum_{j=1}^{\infty} \omega_j \epsilon_{t-j} \tag{3}$$

- E a FACP de um AR(p) estacionário? Pela definição da FACP de ordem k como o coeficiente de Y_{t-k} na regressão populacional de Y_t em Y_{t-1}, Y_{t-2},... Y_{t-k}, esperamos que a FACP seja truncada em p, visto que o processo só depende diretamente das p primeiras defasagens.

Exemplo: FAC e FACP estimadas de um AR(2) (T=10.000)



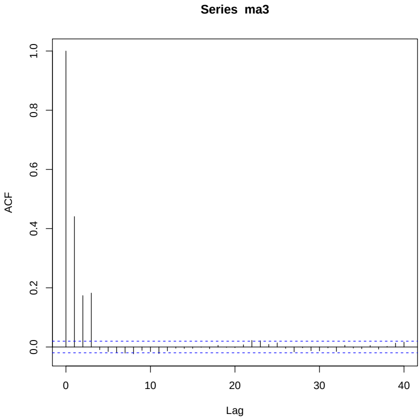


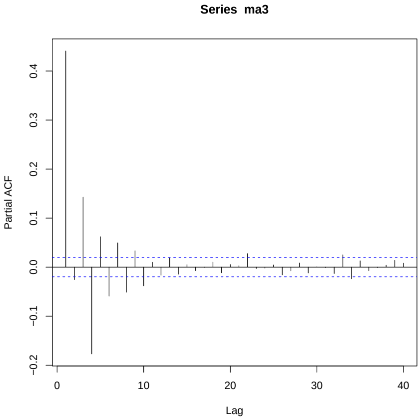
FAC e FACP de um MA(q) invertível

- Como vimos em aula anterior, a FAC de um MA(q) é truncada em q, visto que a correlação morre após q períodos.

- E a FACP? Se o processo MA for invertível, vimos que ele pode ser escrito como um $AR(\infty)$. Dessa representação, fica claro que a FACP de um MA(q) apresenta decaimento em direção a zero.

Exemplo: FAC e FACP estimadas de um MA(3) (T=10.000)

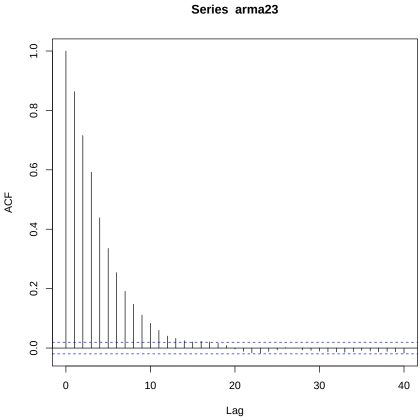


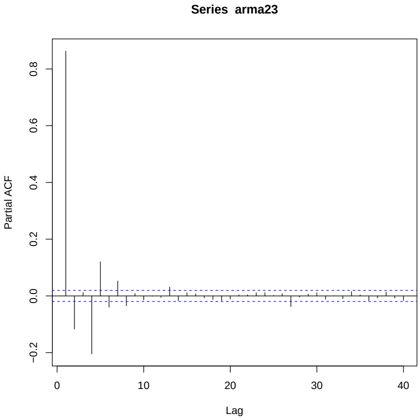


FAC e FACP de um ARMA(p,q) estacionário e invertível

- Generalizando a discussão anterior, um ARMA(p,q) estacionário cuja parte MA é invertível pode ser representado tanto como um $AR(\infty)$ como um $MA(\infty)$. Nesse caso, tanto a FAC como a FACP apresentam decaimento.
- Nesses casos, costuma-se considerar a ordem máxima $q_{\max}$ em que a FAC torna-se pouco significativa e a ordem $p_{\max}$ em que a FACP torna-se pouco signifcativa e considerar todos os ARMA(p,q), $0 \leq p \leq p_{\max}$ e $0 \leq q \leq q_{\max}$ como candidatos.

Exemplo: FAC e FACP estimadas de um ARMA(2,3) (T=10.000)





Resumo

Modelo	FAC	FACP
AR(p) estacionário	decai	truncada em p
MA(q) invertível	truncada em q	decai
ARMA(p,q) estacionário e invertível	decai (esp. após q)	decai (esp. após p)

Estimação

Estimação condicional vs. incondicional

- Para a estimação de modelos ARMA(p,q), há duas abordagens de estimação.
  - Na abordagem condicional, não utilizamos a informação acerca da distribuição das primeiras observações na estimação.
  - Na abordagem incondicional, fazemos hipóteses adicionais sobre a distribuição do ruído branco, que nos permitem incorporar a distribuição das primeiras observações na análise.
- Computacionalmente, a abordagem condicional é mais simples, embora menos eficiente que a segunda.
- Embora a abordagem incondicional aparente requerer mais hipóteses, visto que especificamos a distribuição do ruído branco, a estimação é robusta a violações dessa hipótese quando o número de observações é grande ("pseudo" máxima verossimilhança).

Estimação condicional do AR(p)

conteúdo...

Estimação condicional do ARMA(p,q)

conteúdo...

Estimação incondicional do ARMA(p,q)

conteúdo...