

Trabalho Prático 1

Análise de Séries Temporais - 1/2023



Ana Theresa Figueiredo - 18/0116088

César Augusto Galvão - 19/0011572

Table of contents

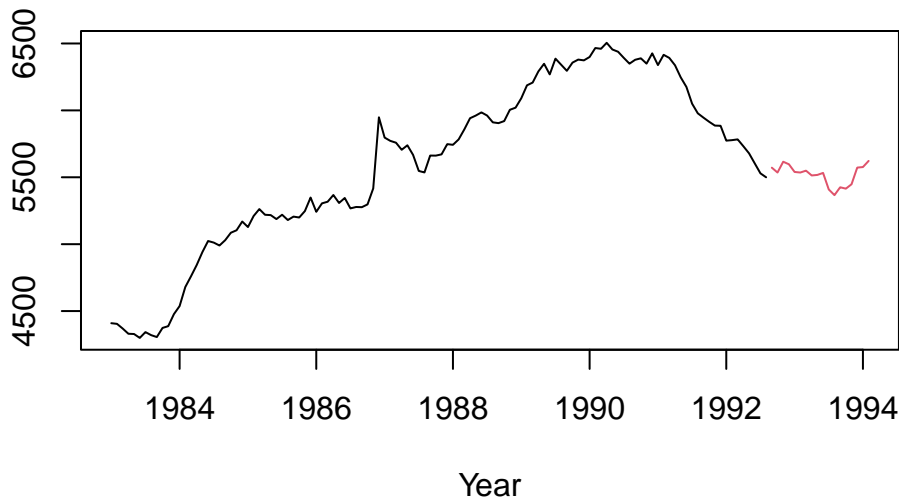
Introdução	2
a. Decomposição da série temporal.	2
b. Escolha um modelo ARIMA	4
Estacionariedade da série	5
Análise gráfica FAC e FACP	5
Análise iterativa	6
c. Análise de resíduos do modelo selecionado.	6
d. Apresentação do modelo selecionado	8
Comparação auto.arima()	8
Apêndice	9

Introdução

A série temporal escolhida foi a de número *id* correspondente a 2546. De acordo com a definição do próprio pacote, refere-se a *Large commercial banks, commercial & industrial loans*, ou seja, dados financeiros e de empréstimos industriais e comerciais. Foram realizadas medidas mensais de 1983 a 1992 e o horizonte de previsão requerido é das 18 ocorrências seguintes.

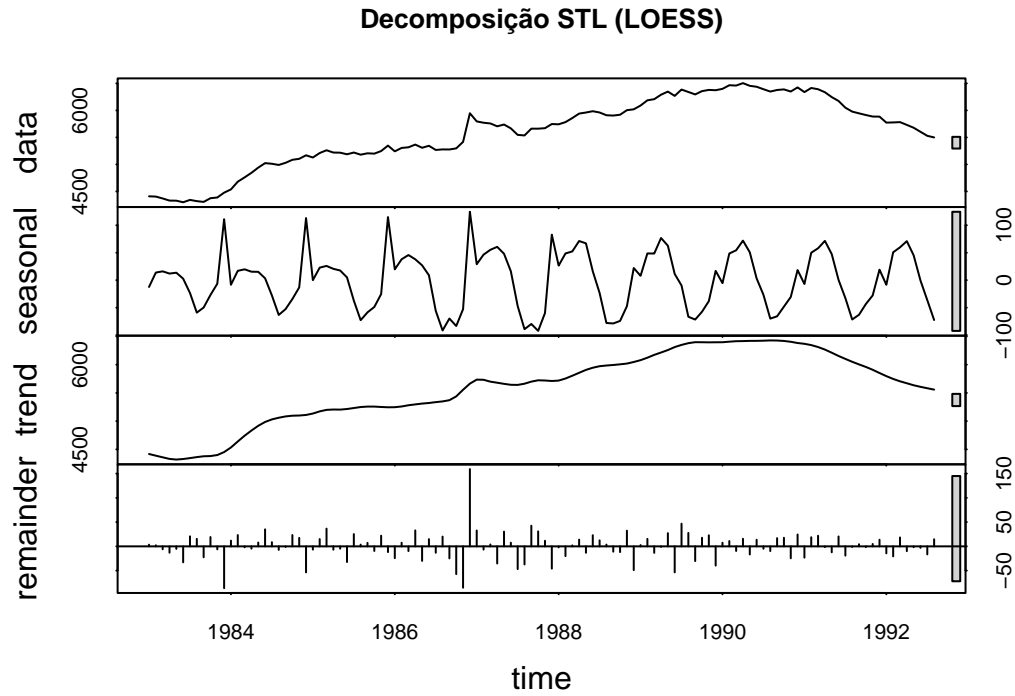
O gráfico da série, com *in* e *out-sample*, é exposto a seguir.

Série Temporal M3–2546



a. Decomposição da série temporal.

Inicia-se a decomposição utilizando STL, caso em que se utiliza uma decomposição Loess. O gráfico dos resultados da função `stl()` são expostos a seguir utilizando os argumentos `s.window` e `t.window` iguais a sete. Esta configuração de argumentos foi escolhida após algumas iterações – balizada também por recomendações de Cleveland *et al.* (1990) – buscando, entre outros fatores, um comportamento adequado dos resíduos do modelo.

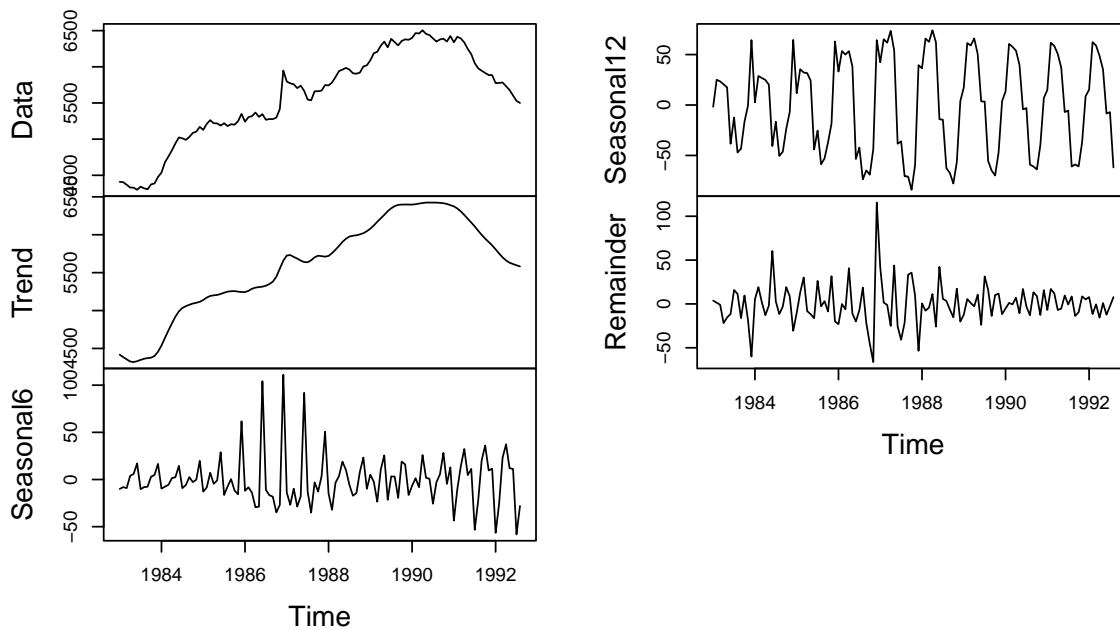


De fato, observa-se uma tendência crescente até 1990, além de parecer haver também uma sazonalidade anual. Esta apresenta alguns picos mais intensos no começo dos ciclos periódicos e menos intensos na segunda metade da série.

Além disso, o termo aleatório parece estável a menos de um pico em 1987, característica que orientou a escolha dos valores para `t.window` e `s.window`.

Experimenta-se também a decomposição via MSTL, procurando uma sazonalidade múltipla, sugerida pelos picos nos inícios das ondas sazonais. Após algumas iterações, optou-se pelo período bimestral, complementarmente ao período mensal original da série. Os mesmos valores para os argumentos de janela foram utilizados.

Decomposição MSTL



De fato, conseguiu-se remover o pico inicial da sazonalidade simples. No entanto, nota-se uma variação grande nas amplitudes sazonais, assim como homocedasticidade do resíduo.

Opta-se finalmente pela decomposição STL, com sazonalidade simples.

b. Escolha um modelo ARIMA

- O modelo selecionado deve ser baseado no que você visualizou na decomposição, testes estatísticos, gráficos ACF, gráficos PACF, critérios de parcimônia e resíduos;
- Detalhe todo o procedimento da escolha do modelo;
- Não pode usar AutoArima
- Tem q fazer o algoritmo mostrado em sala.

Para selecionar o modelo ARIMA, iniciamos pela identificação das diferenciações sazonais e não sazonais utilizando as funções `nsdiffs()` e `ndiffs()`, respectivamente, do pacote `forecast`. Dessa forma, para um modelo $SARIMA(p, d, q) \times (P, D, Q)_{12}$ ou $SARMA(P, Q)_{12}$, obtemos os valores de D e d .

Isto feito, sabemos que o modelo é pelo menos $(p, 2, q) \times (P, 0, Q)_{12}$.

Estacionariedade da série

Antes de prosseguir, avalia-se a estacionariedade da série. Para isso, utiliza-se o teste KPSS (Kwiatkowski *et al.*, 1992). As hipóteses do teste são:

H_0 : o processo é estacionário

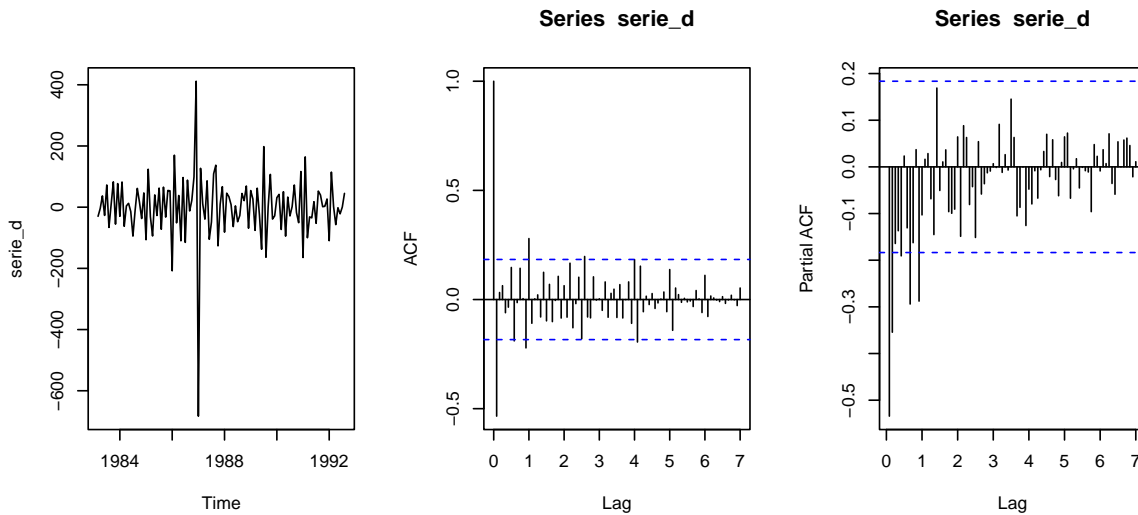
H_1 : o processo possui raiz unitária

Os resultados do teste são exibidos na tabela a seguir, sugerindo não rejeitar a hipótese de estacionariedade.

Estatística de teste	p-valor	Método
0.0244372	0.1	KPSS Test for Level Stationarity

Análise gráfica FAC e FACP

A seguir são expostos três gráficos para a série diferenciada em $\nabla^2 x_t$: a série original com duas diferenciações simples (portanto estacionária), sua Função de Autocorrelação - FAC (*ACF*) e a Função de Autocorrelação Parcial - FACP (*Partial ACF*).



Nota-se um decaimento amortecido no gráfico FAC e poucas correlações fora da banda apenas até Lag 1 do FACP. As autocorrelações parciais, no entanto, também decaem de forma amortecida para zero sem quebras. Este comportamento indica candidatos a modelos ARIMA ou SARMA.

Análise iterativa

Na falta de comportamentos claros indicando os demais coeficientes, recorre-se a uma varredura para selecionar o modelo mais parcimonioso utilizando AICc como critério.

Conforme a saída a seguir, verifica-se uma boa configuração sendo SARMA(0, 2, 1) \times (1, 0, 0)



```
[1] "p = 0, q = 1, P = 1, Q = 0, AICc = 1313.1422"
[2] "p = 0, q = 1, P = 0, Q = 1, AICc = 1313.8014"
[3] "p = 0, q = 1, P = 0, Q = 0, AICc = 1318.3233"
[4] "p = 0, q = 0, P = 1, Q = 0, AICc = 1380.2924"
[5] "p = 0, q = 0, P = 0, Q = 1, AICc = 1380.984"
[6] "p = 0, q = 0, P = 0, Q = 0, AICc = 1387.0622"
```

Finalmente, ajusta-se o modelo conforme as especificações e obtém-se os seguintes coeficientes:

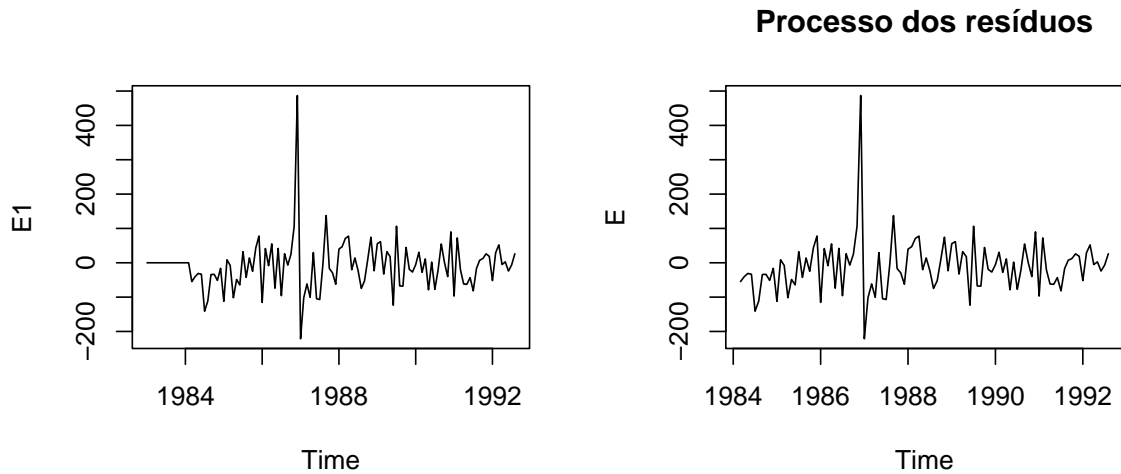
Coeficiente	Valor
ma1	-0.8394
sar1	0.2545

c. Análise de resíduos do modelo selecionado.

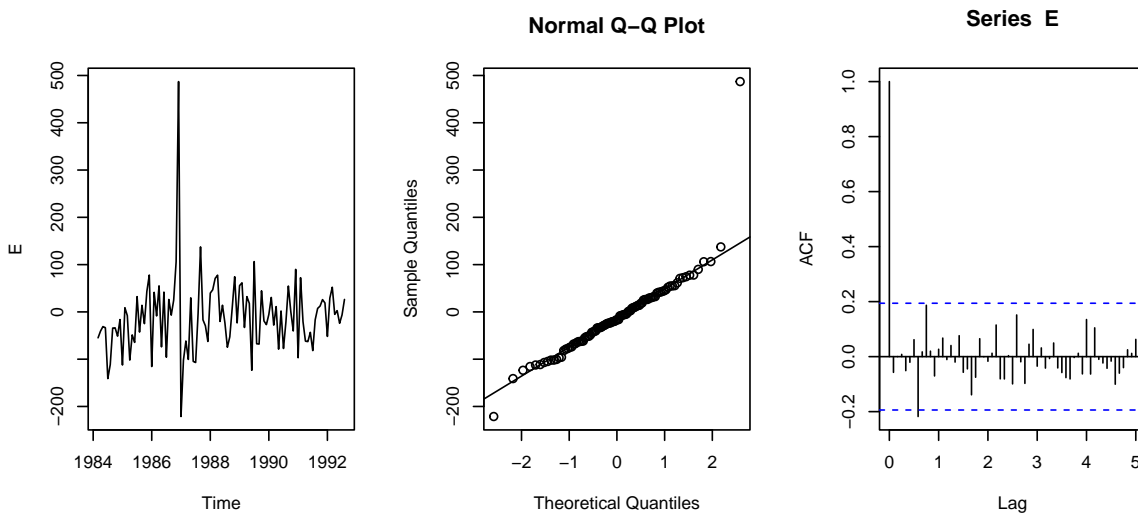
Para avaliar a qualidade do modelo, verifica-se as seguintes suposições sobre $\{\varepsilon_t\}$:

- Média zero,
- Variância constante,
- Autocorrelação nula,
- Normalidade.

Verifica-se inicialmente a série dos resíduos sem a inicialização com zeros. Graficamente, pode-se dizer que tem média igual a zero. De fato, a média calculada dos resíduos corresponde a -9.495, o que é muito próximo de zero considerando a ordem de grandeza de muitos valores dos resíduos. Além disso, a variância parece ser constante, a menos do pico no meio da série.



O Q-Q Plot a seguir aponta uma proximidade não ideal em relação à distribuição normal com uma cauda mais pesada à esquerda e um ponto extremo à direita. Além disso, o ACF possui uma quebra na banda de confiança, mas mantém as demais autocorrelações dentro dos limites.



Por fim, verifica-se utilizando testes de hipótese estacionariedade (H_0 : o processo é estacionário), independência (H_0 : $\rho(1) = \rho(2) = \dots = \rho(m) = 0$) e normalidade (H_0 : não se rejeita normalidade) dos resíduos do modelo.

Infere-se pelos testes que se trata de uma série estacionária, com autocorrelações todas iguais a zero, porém não normal. Isso provavelmente ocorre por causa de um peso maior na cauda esquerda e por causa da presença de alguns valores extremos.



Estatística de teste	p-valor	Método
0.2061	0.1000	KPSS Test for Level Stationarity
15.7901	0.7296	Box-Ljung test
0.8363	0.0000	Shapiro-Wilk normality test



d. Apresentação do modelo selecionado

O modelo selecionado é, conforme já apresentado, $\text{SARMA}(0, 2, 1) \times (1, 0, 0)_{12}$. Em forma polinomial temos:



$$\begin{aligned}
 \Phi_P(B^s) \nabla^2 x_t &= \Theta(B) \varepsilon_t \\
 \Phi_1(B) \nabla^2 x_t &= \Theta_1(B) \varepsilon_t \\
 (1 - \phi_1 B)(x_t - 2x_{t-1} + x_{t-2}) &= (1 + \theta_1 B) \varepsilon_t \\
 x_t - 2x_{t-1} + x_{t-2} - \phi_1 x_{t-1} - 2\phi_1 x_{t-2} - \phi_1 x_{t-3} &= \varepsilon_t + \theta_1 \varepsilon_{t-1} \\
 x_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + 2x_{t-1} - x_{t-2} + \phi_1 x_{t-1} + 2\phi_1 x_{t-2} + \phi_1 x_{t-3}
 \end{aligned}$$

Finalmente, obtemos

$$\begin{aligned}
 x_t &= \varepsilon_t - 0.839 \varepsilon_{t-1} + 2x_{t-1} - x_{t-2} + 0.255 x_{t-1} + 0.51 x_{t-2} + 0.255 x_{t-3} \\
 &= \varepsilon_t - 0.839 \varepsilon_{t-1} + 2.255 x_{t-1} - 0.49 x_{t-2} + 0.255 x_{t-3}
 \end{aligned}$$

Comparação `auto.arima()`

Por fim, compara-se o modelo com a saída do modelo sugerido pela função automatizada. De fato, são próximos e o modelo é o mesmo que o selecionado manualmente.

```

Series: serie$x
ARIMA(0,2,1)(1,0,0)[12]

Coefficients:
      ma1      sar1
    -0.9385    0.2455
s.e.    0.0404    0.0891

sigma^2 = 5547:  log likelihood = -653.46
AIC=1312.92   AICc=1313.14   BIC=1321.13

```


Apêndice

Todo o projeto de composição deste documento pode ser encontrado aqui: https://github.com/cesar-galvao/trabalhos_series

```
pacman::p_load(Mcomp, tidyverse, forecast, fpp2, xts, tseries, tidymodels)

data(M3) #carrega os dados
id <- 2546 #série temporal escolhida

serie <- M3[[id]]

dados <- serie$x

plot(serie, main = "Série Temporal M3-2546")

serie$x %>%
  stl(s.window = 7, t.window = 7) %>%
  plot(main = "Decomposição STL (LOESS)")

#sazonalidade diaria e semanal
msts(serie$x, seasonal.periods = c(12, 6))%>%
#janela de ajuste 'Cleveland et al (1990)'
  mstl(., s.window = 7, t.window = 7) %>% plot(main = "Decomposição MSTL")

d <- ndiffs(serie$x)

D <- serie$x %>% diff() %>% nsdiffs()

#aplica 2 diferenciacoas
serie_d <- serie$x %>% diff(differences = 2)

kpss.test(serie_d)

par(mfrow = c(1, 3))
plot(serie_d)
acf(serie_d, lag.max = 12*7)
pacf(serie_d, lag.max = 12*7)

parcim <- character()
```

```

melhor_AICc <- Inf
for(p in 0:2){
  for(q in 0:2){
    for(P in 0:1){
      for(Q in 0:1){
        fit <- Arima(serie$x, order = c(p,2,q), seasonal = c(P, 0, Q))
        if(fit$aicc < melhor_AICc){
          melhor_AICc <- fit$aicc
          parcim <- c(paste0("p = ",p," ", q = " ", q, " ", P = " ", P, " ", Q = " ", Q,
                                " ", AICc = " ", round(fit$aicc,4)),parcim)
        }
      }
    }
  }
}

```

parcim

```

fit <- arima(serie$x, order = c(0,2,1), seasonal = c(1,0,0), method = "CSS")

```

```

fit$coef %>% tidy()

```

```

par(mfrow=c(1,2))
E <- fit$residuals %>% window(start=c(1984,3))
E1 <- fit$residuals
plot(E1)
plot(E, main = "Processo dos resíduos")

```

```

par(mfrow=c(1,3))
plot(E)
qqnorm(E)
qqline(E)
acf(E, lag.max=12*5)

```

```

bind_rows(
  tseries::kpss.test(E) %>% tidy() %>% select(-parameter),

```

```
Box.test(E, lag = 20, type = "Ljung-Box")%>% tidy() %>% select(-parameter)

auto.arima(serie$x, d = 2, D = 0)
```