Identificação de Modelos ARIMA

Lucas Resck e Lucas Moschen

October 13, 2020

1. (Questão 30 - 5) Considere o modelo abaixo.

$$Z_t = \sum_{j=0}^{m} \beta_j t^j + \frac{\theta(B)}{\phi(B)\Delta^d} a_t$$

Prove que, se $m > d$:

a. Tomando-se d diferenças, obtemos um modelo não estacionário, com uma tendência polinomial de grau $m - d = h$.

Temos que $\Delta^d Z_t = (1-B)^d Z_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k B^k Z_t$, usando o binômio de Newton. Em particular:

$$\begin{aligned}
\Delta^d t^j &= \sum_{k=0}^{d} \binom{d}{k} (-1)^k B^k t^j \\
&= \sum_{k=0}^{d} \binom{d}{k} (-1)^k (t-k)^j \\
&= \sum_{k=0}^{d} \binom{d}{k} (-1)^k \sum_{i=0}^{j} \binom{j}{i} t^i k^{j-i} (-1)^{j-i} \\
&= \sum_{i=0}^{j} \binom{j}{i} t^i (-1)^{j-i} \sum_{k=0}^{d} \binom{d}{k} (-1)^{d-k} (d-k)^{j-i}, \text{ redefinindo } k = d-k \\
&= \sum_{i=0}^{j} \binom{j}{i} t^i (-1)^{j-i} \sum_{k=0}^{d} \binom{d}{k} (-1)^{d+k} (d-k)^{j-i} \\
&= \sum_{i=0}^{j} \binom{j}{i} t^i (-1)^{j-i} (-1)^d \sum_{k=0}^{d} \binom{d}{k} (-1)^{k} (d-k)^{j-i} \\
&= \sum_{i=0}^{j} \binom{j}{i} t^i (-1)^{j-i} d! (-1)^d \left\{ {j-i \atop d} \right\}, \text{ onde o último é o número de Stirling de 2º tipo}
\end{aligned}$$

Sabemos que se $d > j - i \implies i > j - d$, o número de Stirling é 0, isto é, $\Delta^d t^j$ é um polinômio de ordem $(j-d)^+$. Em particular, se $d > j$, teremos um polinômio de ordem 0.

$$\Delta^d Z_t = \sum_{j=0}^{m} \beta_j \Delta^d t^j + \frac{\theta(B)}{\phi(B)} a_t$$

Defina $m - d = h > 0$. Quando $j = m$, teremos que $\Delta^d t^m$ será um polinômio de ordem h, pelo que vimos acima, dado que o número de Stirling é diferente de 0.

b. Tomando-se m diferenças obteremos um processo estacionário não invertível.

Tomando m diferenças, teremos um polinômio de ordem 0. Seja, então, se olharmos a forma acima, o único valor diferente de 0 será quando $i = 0, j = d = m$.

$$\Delta^m Z_t = \beta_m m! + \frac{\theta(B)}{\phi(B)} a_t$$

Esse processo não é invertível porque não podemos escrever $a_t = \pi(B) Z_t$, dado à presença do nível.

2. (Questão 31 - 5) Prove que se $W_t = (1 - B) Z_t$, então $Z_t = W_t + W_{t-1} + ...$

Podemos representar $\frac{1}{1-B}$ como:

$$\frac{1}{1-B} = \sum_{j=0}^{\infty} B^j,$$

portanto $Z_t = \sum_{j=0}^{\infty} B^j W_t = W_t + W_{t-1} + W_{t-2} + ...$

3. (Questão 32 - 5) Prove que, na forma invertida do modelo, $\sum_{j=1}^{\infty} \pi_j = 1$.

Na forma invertida do modelo, ele é escrito da seguinte forma:

$$\pi(B) Z_t = a_t$$

Vamos expandir o polinômio e tomar o valor esperado:

$$
\begin{aligned}
a_t &= \left[1 - \sum_{j=1}^{\infty} \pi_j B^j\right] Z_t \\
a_t &= Z_t - \pi_1 Z_{t-1} - \pi_2 Z_{t-2} - \cdots \\
\mathbb{E}\{a_t\} &= \mathbb{E}\{Z_t - \pi_1 Z_{t-1} - \pi_2 Z_{t-2} - \cdots\} \\
0 &= \left(1 - \sum_{j=1}^{\infty} \pi_j\right) \mu \\
0 &= 1 - \sum_{j=1}^{\infty} \pi_j \\
\sum_{j=1}^{\infty} \pi_j &= 1
\end{aligned}
$$

Observe que assumimos que $\mu \neq 0$. Isso na verdade é sem perda de generalidade, afinal:

$$
\begin{aligned}
\pi(B)Z_t &= a_t \\
\theta(B)\pi(B)Z_t &= \theta(B)a_t \\
\varphi(B)Z_t &= \theta(B)a_t \\
\phi(B)\Delta^d Z_t &= \theta(B)a_t \\
\phi(B)\Delta^d (Z_t + \alpha) &= \theta(B)a_t \\
\phi(B)\Delta^d Y_t &= \theta(B)a_t \\
&\vdots \\
\pi(B)Y_t &= a_t \\
\pi(B)(Z_t + \alpha) &= a_t
\end{aligned}
$$

5. (Questão 1 - 6) Prove que se $\rho_j = \phi^{|j|}, |\phi| < 1$, então

$$
Var(r_j) = \frac{1}{N}\left[\frac{(1+\phi^2)(1-\phi^{2j})}{1-\phi^2} - 2j\phi^{2j}\right],
$$

em particular $Var(r_1) = \frac{1}{N}(1-\phi^2)$.

Observe que, se vale que $\rho_j = \phi^{|j|}$, então essas correlações são do modelo AR(1), como visto em aula:

$$
Z_t = \phi Z_{t-1} + a_t
$$

Ora, Fuller (2009) mostrou que, nesse caso,

$$\mathrm{Var}(r_j) = \frac{N-j}{N^2}\left[\frac{(1+\phi^2)(1-\phi^{2j})}{1-\phi^2} - 2j\phi^{2j}\right]$$

Acreditamos que a demonstração está fora do escopo deste exercício, pois envolve vários teoremas e bastante manipulações algébricas. Em particular:

$$\begin{aligned}
\mathrm{Var}(r_1) &= \frac{N-1}{N^2}\left[\frac{(1+\phi^2)(1-\phi^2)}{1-\phi^2} - 2\phi^2\right] \\
&= \frac{N-1}{N^2}\left[\frac{1-\phi^4}{1-\phi^2} - 2\phi^2\right] \\
&= \frac{N-1}{N^2}\left[\frac{1-\phi^4-2\phi^2+2\phi^4}{1-\phi^2}\right] \\
&= \frac{N-1}{N^2}\left[\frac{(1-\phi^2)^2}{1-\phi^2}\right] \\
&= \frac{N-1}{N^2}(1-\phi^2)
\end{aligned}$$

---

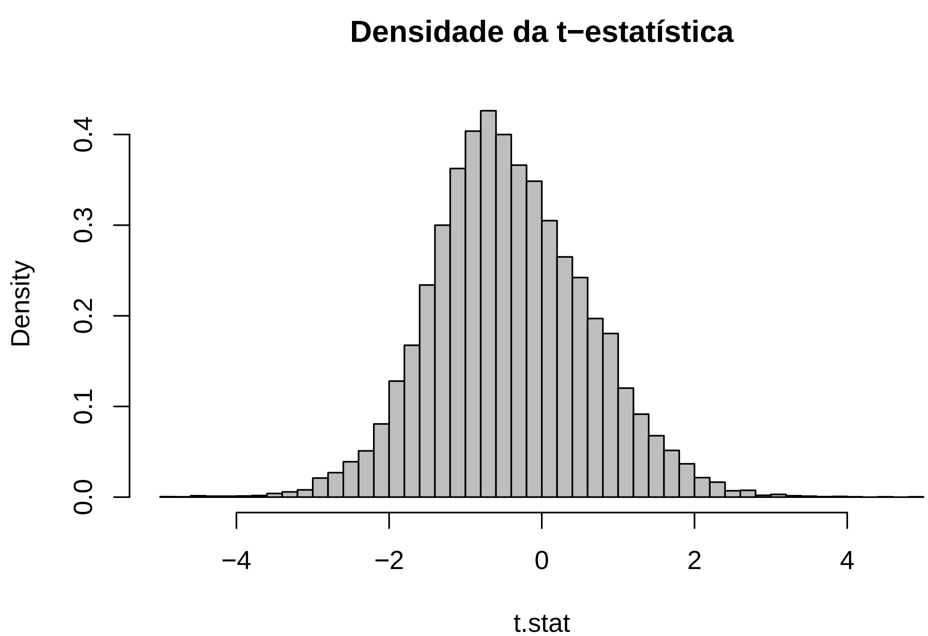
2. Simulação da distribuição da estatística de teste de Dickey-Fuller Acesse o livro 'Econometria de Séries Temporais' 2a. Edição do Rodrigo de Losso Bueno na "Minha Biblioteca" Leia Cap 4.5.1, 4.5.2 e 4.5.3, Pag 116 (ou 134 online)

```r
df.simulation <- function(n, Tr, S, e.sd, phi, intercept = 0){
  t.stat <- c(1:S)
  for(i in 1:S){
    # Passo 1
    e <- rnorm(n + Tr, mean = 0, sd = e.sd)
    # Passo 2
    y <- e + intercept
    for(j in (length(phi)+1):(n+Tr)){
      for(k in 1:length(phi)){
        y[j] <- phi[k]*y[(j-k)] + y[j]
      }
    }
    # Passo 3 e 4
    x <- y[(n+1):(n+Tr)]
    mom2 <- sum(x[1:(Tr-1)]*x[1:(Tr-1)])
    # Us
    model <- ar.ols(x, aic = F, order.max = length(phi),
                    intercept = !(intercept==0), demean = F)
    rho.hat <- sum(model$ar)
```

```r
    error <- x[(length(phi)+1):Tr] - intercept
    for(k in length(phi)){
      error <- error - model$ar[[k]]*x[(length(phi) - k + 1):(Tr - k)]
    }
    Sn <- sum(error^2)/(Tr-(length(phi)+1))

    t.stat[i] = (rho.hat - 1)*sqrt(mom2)/Sn
  }
  return(t.stat)
}
```

Primeiro, para fins de comparação, vejamos a distribuição quando $\phi = 1$.

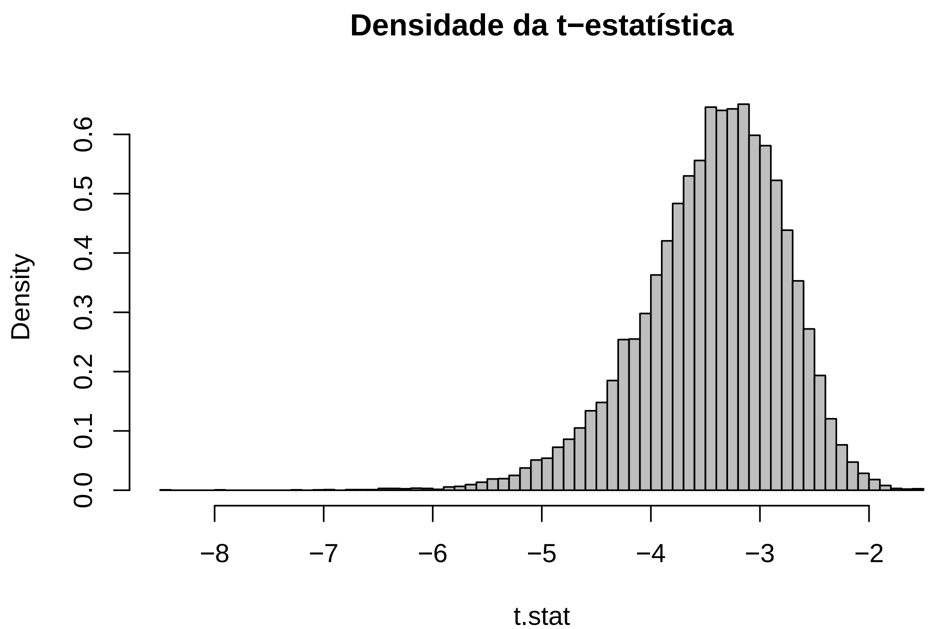
```r
set.seed(10000)

n <- 50
Tr <- 100
S <- 20000
e.sd <- 1
phi <- c(1)

t.stat <- df.simulation(n,Tr,S,e.sd,phi)
hist(t.stat, breaks = 50 , freq = F, col = 'grey',
      main = 'Densidade da t-estatística')
```



Vemos que o gráfico é similar com o exemplo do livro, o que é um *sanity check* interessante.

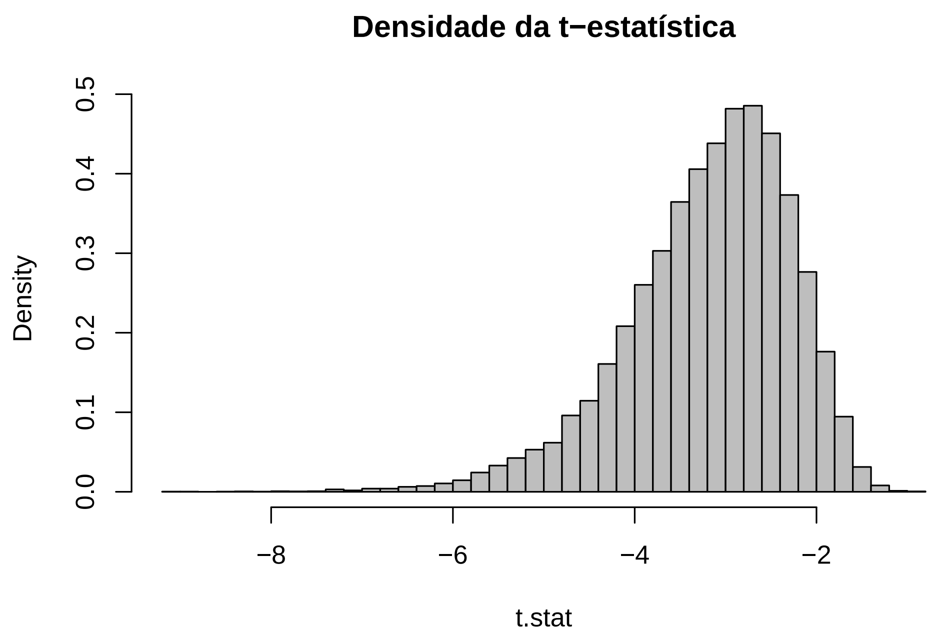
2.1 Simule os valores críticos da estatística de teste DF, como proposto no Cap 4.5.1, após eq (2) com $\phi_1 = 0.8$.

```
phi <- c(0.8)

t.stat <- df.simulation(n,Tr,S,e.sd,phi)
hist(t.stat, breaks = 50 , freq = F, col = 'grey',
      main = 'Densidade da t-estatística')
```



2.2 Repita a simulação, mas agora adicione termos autoregressivos no modelo, verifique que a distribuição da estatística de teste permanece inalterada.

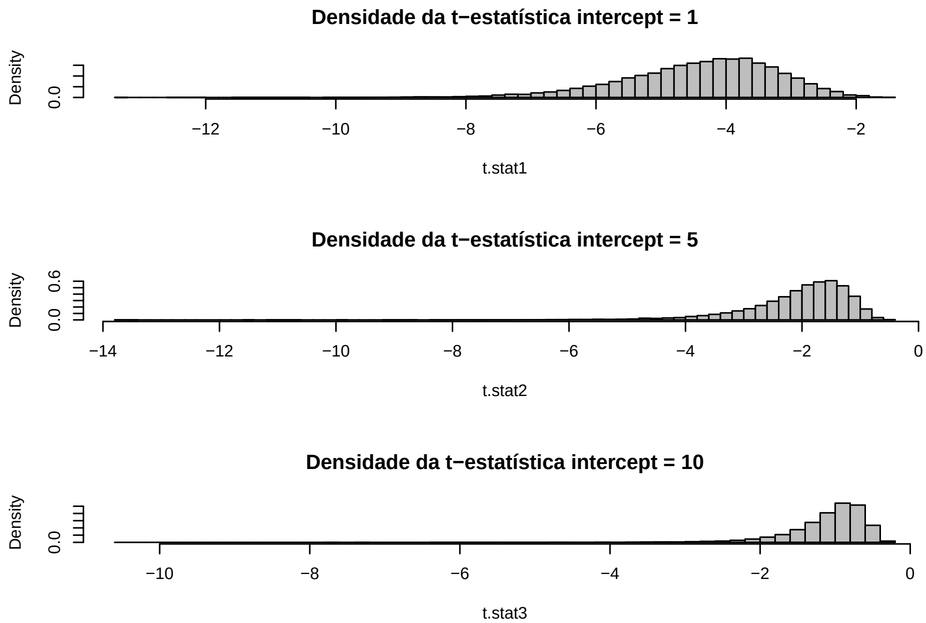
```
phi <- c(0.8,-0.4, 0.2)

t.stat <- df.simulation(n,Tr,S,e.sd,phi)
hist(t.stat, breaks = 50 , freq = F, col = 'grey',
      main = 'Densidade da t-estatística', )
```



De fato as distribuições são bem similares, como podemos observar pelos gráficos. Isso é provado no Apêndice B do livro de referência.

2.3 Adicione intercepto e verifique se a nova distribuição da estatística de teste muda.

```
phi <- c(0.8, -0.4, 0.2)

t.stat1 <- df.simulation(n,Tr,S,e.sd,phi, intercept = 1)
t.stat2 <- df.simulation(n,Tr,S,e.sd,phi, intercept = 5)
t.stat3 <- df.simulation(n,Tr,S,e.sd,phi, intercept = 10)
```

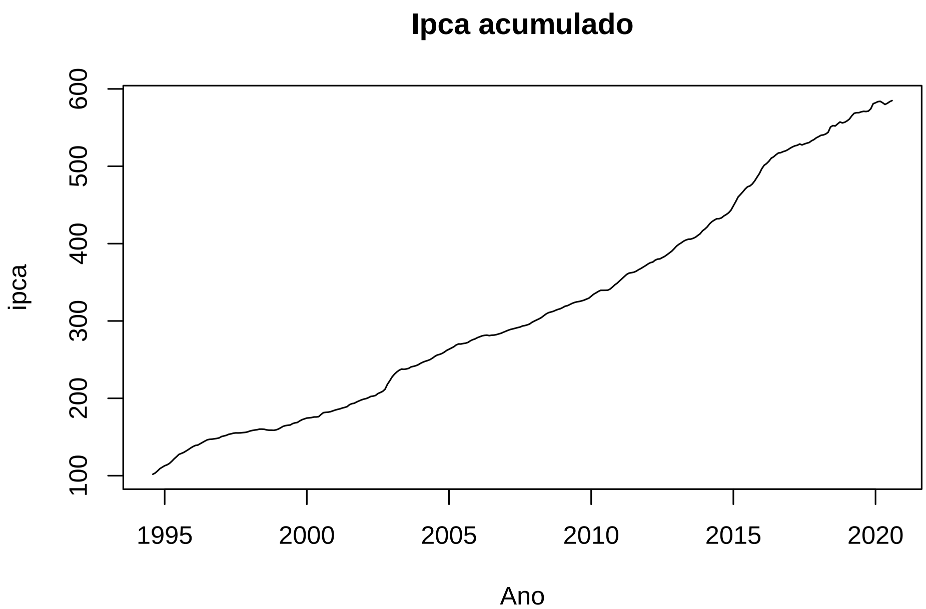


Vemos que a distribuição varia conforme o intercepto. Quando o intercepto é pequeno, a distribuição parece não mudar muito, porém quando ele aumenta, o histograma vai mudando um pouco de formato, concentrando-se mais e mais próximo de 0.

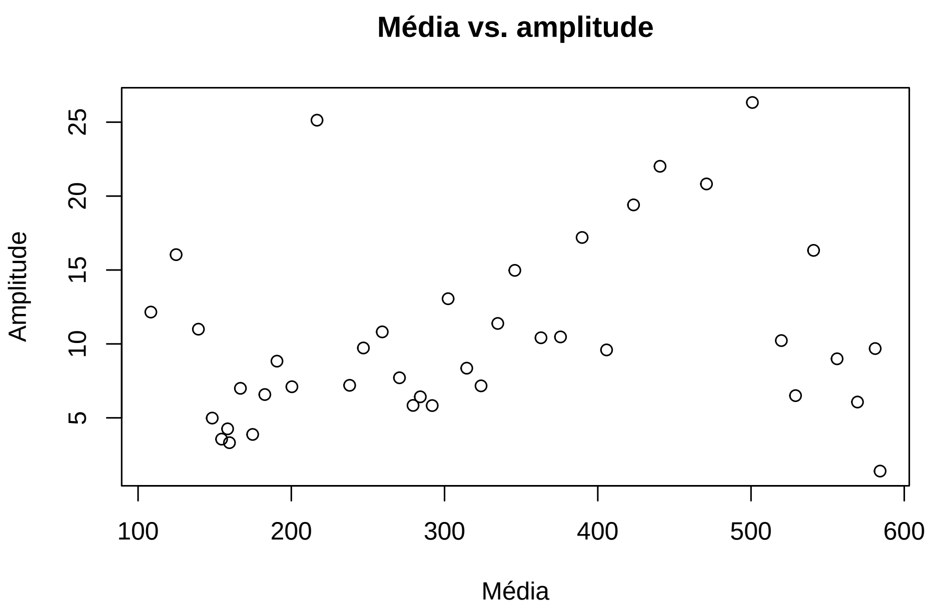
---

3. Identificação do modelo: Identifique o modelo ARIMA para a série de dados de inflação

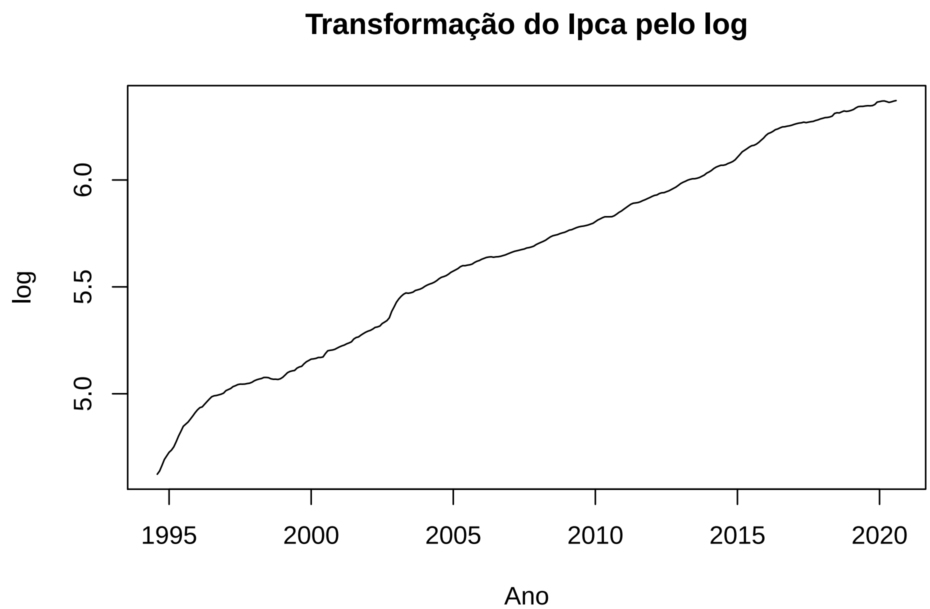
Primeiro podemos verificar com que tipo de série estamos lidando através de um plot do gráfico. Os dados são de caráter mensal e se inicia em agosto de 1994 até agosto de 2020. Não há dados faltantes nem duplicados.



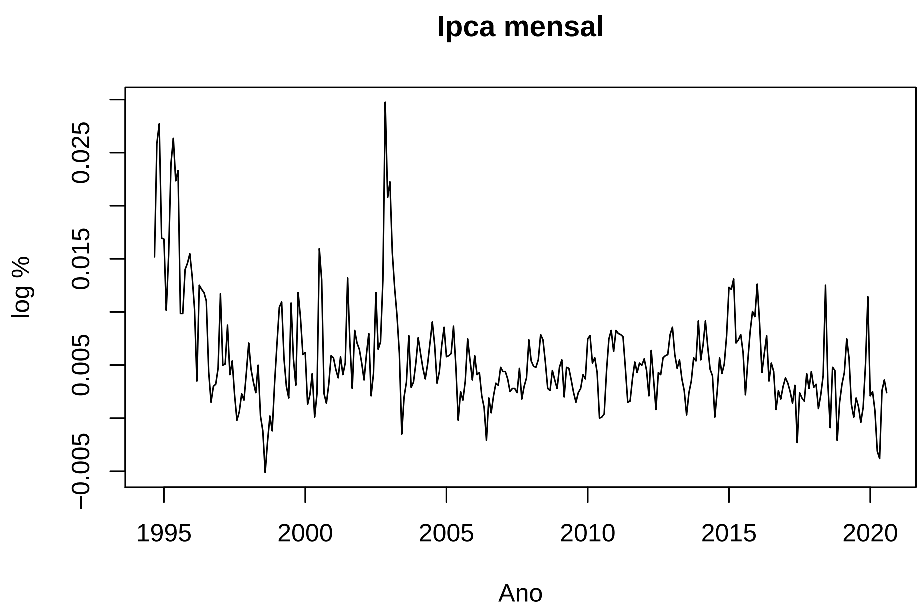
Vejamos se é necessário uma transformação para estabilizar a variância. O que faremos é dividir a nossa série temporal em um conjunto de 8 observações consecutivas, e estimar a média e a amplitude de cada um.



É difícil de tirar grandes conclusões a partir desse gráfico, mas podemos supor que a média cresce linearmente com a amplitude. Sendo assim, sabemos que a transformação indicada é a logarítmica.



Está claro que essa série tem uma tendência, pelo fato de estarmos tomando ipca acumulado. Vamos capturar o valor mensal, portanto:

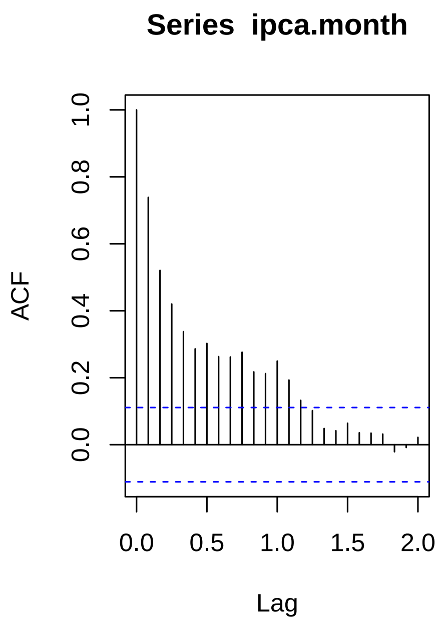


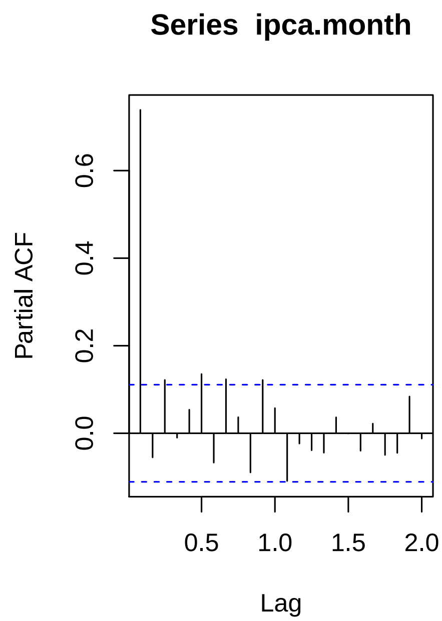
Checando o Teste Dickey-Fuller aumentado, temos que o p-valor é menor do que 0.01, o que faz com que rejeitemos a hipótese de que a série é não estacionária. Concluímos que esse modelo já pode ser utilizado para identificar os parâmetros p e q.

```r
tseries::adf.test(ipca.month, alternative = "stationary")
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ipca.month
## Dickey-Fuller = -4.8786, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

Antes disso, vejamos a ACF e a PACF





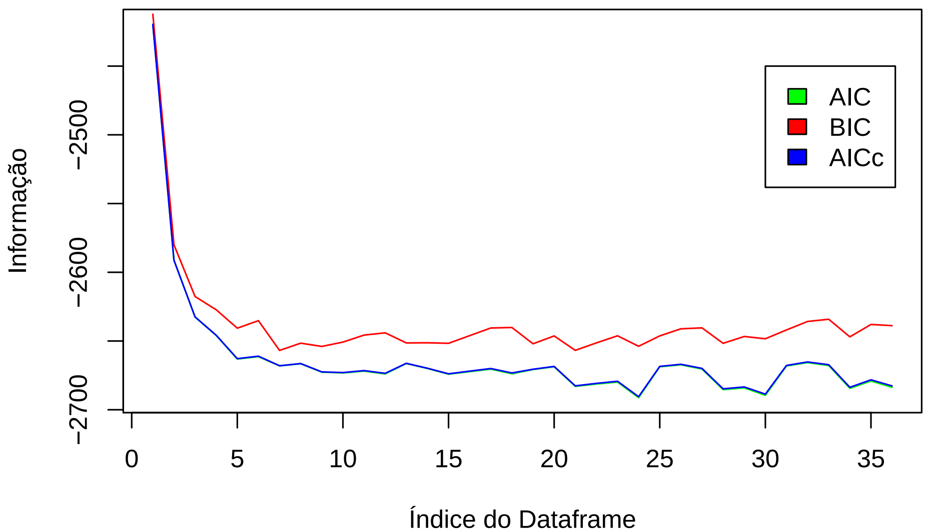
Olhando o gráfico fica difícil discernir. Mas é aparente que a PACF morre logo nos primeiros valores. Portanto, apenas olhando os gráficos, chegaríamos em um modelo AR(p). Em contrapartida, vamos ver como se comportam os critérios de informação. Vamos obter em um dataframe o AIC, BIC e AICC. Estamos tomando $K = L = log(N)$ como valor máximo.

p	q	AIC	BIC	AICC
0	0	-2419.676	-2412.190	-2419.638
0	1	-2591.263	-2580.034	-2591.185
0	2	-2632.593	-2617.620	-2632.462
0	3	-2645.967	-2627.252	-2645.771
0	4	-2663.121	-2640.663	-2662.845
0	5	-2661.393	-2635.192	-2661.025
1	0	-2668.081	-2656.852	-2668.003
1	1	-2666.535	-2651.563	-2666.404

p	q	AIC	BIC	AICC
1	2	-2672.661	-2653.946	-2672.465
1	3	-2673.231	-2650.773	-2672.955
1	4	-2671.893	-2645.692	-2671.524
1	5	-2673.986	-2644.042	-2673.511
2	0	-2666.359	-2651.387	-2666.229
2	1	-2669.995	-2651.280	-2669.799
2	2	-2674.126	-2651.668	-2673.850
2	3	-2672.288	-2646.087	-2671.919
2	4	-2670.485	-2640.541	-2670.009
2	5	-2673.840	-2640.153	-2673.244
3	0	-2670.728	-2652.013	-2670.532
3	1	-2668.763	-2646.305	-2668.488
3	2	-2683.002	-2656.801	-2682.633
3	3	-2681.298	-2651.354	-2680.823
3	4	-2679.882	-2646.195	-2679.286
3	5	-2691.250	-2653.820	-2690.519
4	0	-2668.739	-2646.281	-2668.464
4	1	-2667.310	-2641.109	-2666.942
4	2	-2670.393	-2640.449	-2669.918
4	3	-2685.374	-2651.687	-2684.778
4	4	-2684.140	-2646.710	-2683.409
4	5	-2689.527	-2648.354	-2688.647
5	0	-2668.143	-2641.942	-2667.775
5	1	-2665.715	-2635.771	-2665.240
5	2	-2667.857	-2634.170	-2667.261
5	3	-2684.414	-2646.984	-2683.683
5	4	-2679.120	-2637.947	-2678.240
5	5	-2683.735	-2638.819	-2682.692

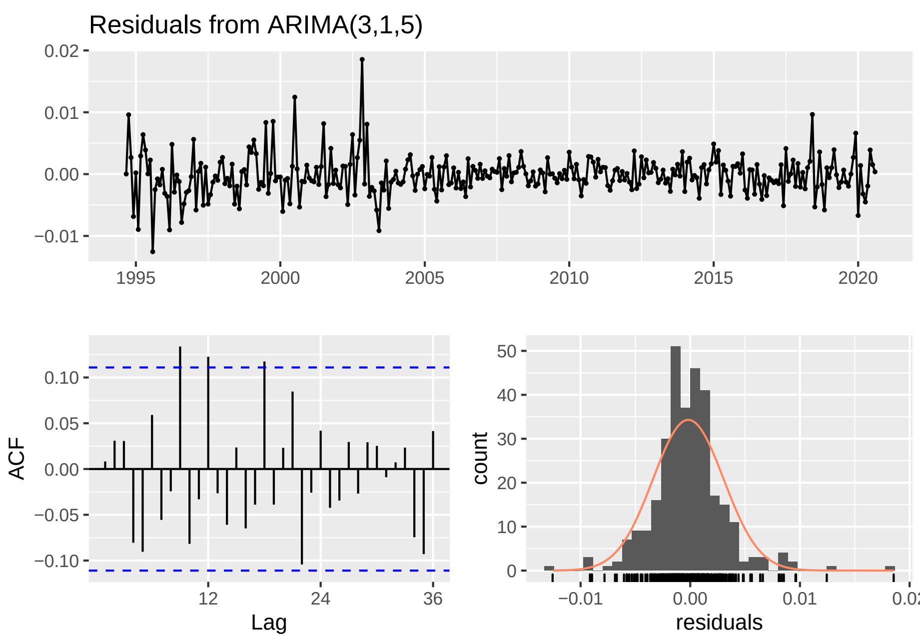
Os critérios AIC e AICc tem informação mínima quandp $p = 3, q = 5$, o modelo com ordem alta (como já era esperado). Enquanto isso, o BIC subestima a ordem do modelo com $p = 1, q = 0$, o que corrobora os gráficos ACF e PACF.

O seguinte gráfico apenas ajuda a identificar quais os índices do dataframe que contêm os parâmetros que minimizam as informações.



Agora nós vamos ajustar os dois modelos encontrados aos dados e analisar os resíduos.

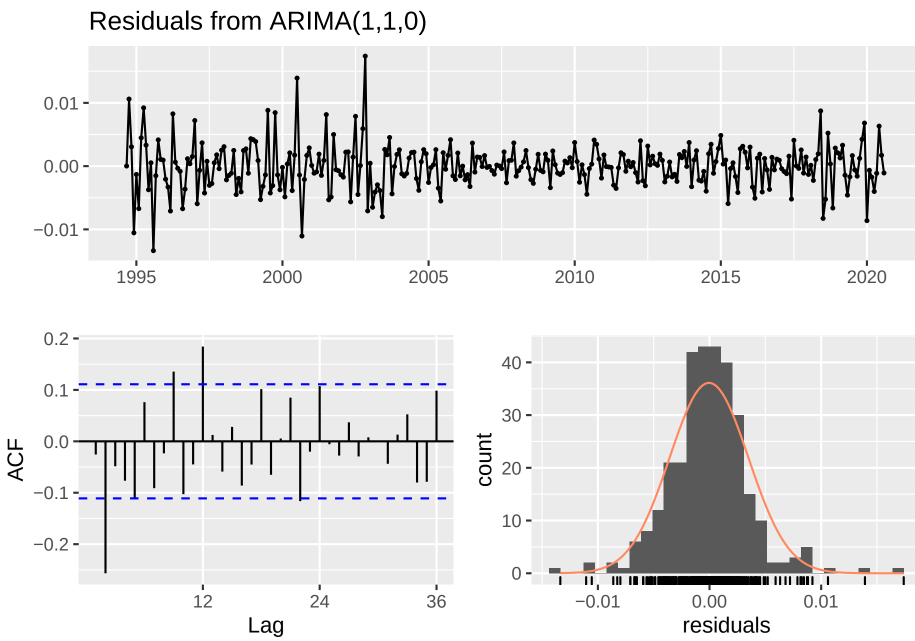
```
model1 = arima(ipca.month, order = c(3, 1, 5))
checkresiduals(model1)
```




```
##
##  Ljung-Box test
##
```

```
## data:  Residuals from ARIMA(3,1,5)
## Q* = 36.547, df = 16, p-value = 0.002428
##
## Model df: 8.   Total lags used: 24
```

```r
model2 = arima(ipca.month, order = c(1, 1, 0))
checkresiduals(model2)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,0)
## Q* = 73.982, df = 23, p-value = 2.908e-07
##
## Model df: 1.   Total lags used: 24
```

A partir dos gráficos, podemos concluir que os resíduos não são normais. Os resultados do teste de Ljung-Box indicam que ambos os modelos não possuem resíduos não correlacionados. Porém, o p-valor do modelo ARIMA(1, 1, 0) é menor, o que indica um melhor ajuste do modelo ARIMA(3, 1, 5). Vamos calcular o MAPE de cada um dos modelos ajustados:

```
## [1] "MSE do ARIMA(3, 1, 5): " "1.03957837630611e-05"
```

```
## [1] "MSE do ARIMA(1, 1, 0): " "1.24579053791551e-05"
```

Observamos que o MSE do ARIMA(3, 1, 5) é melhor. Portanto, escolhemos este modelo para nossos dados.

Referências

Fuller, Wayne A. 2009. *Introduction to Statistical Time Series*. Vol. 428. John Wiley & Sons.