

Modelagem ARIMA e Análise de resíduos

Lucas Resck e Lucas Moschen

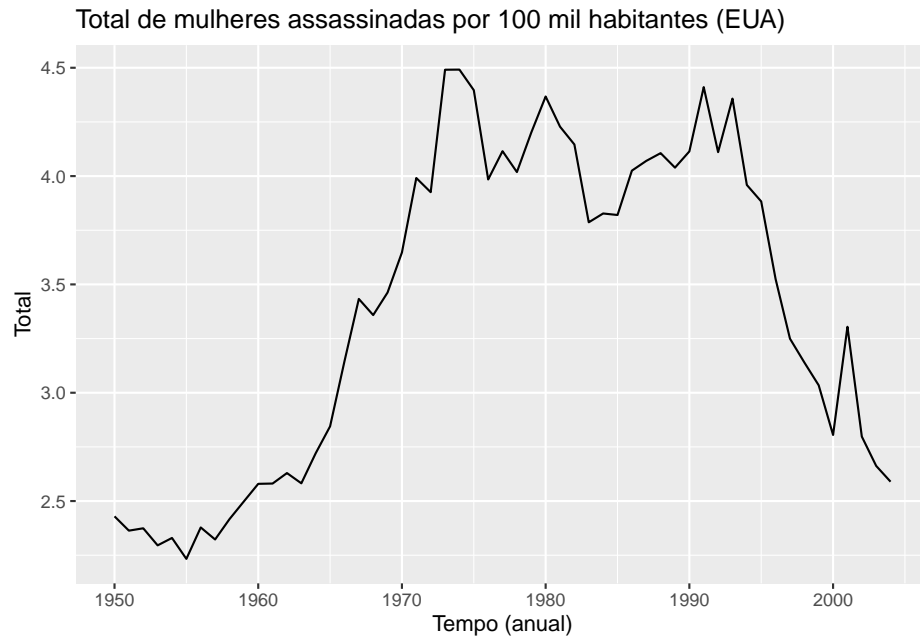
November 6, 2020

Método Box-Jenkins

0. Tranformação dos dados para estabilizar a variância.
 1. Identificação
 - 1.1 Checar a estacionaridade e diferencial d vezes;
 - 1.2 Visualizar autocorrelação e autocorrelação parcial dos dados;
 - 1.3 Comparar informações AIC, BIC e AICc e seleccionar p e q .
 2. Estimação
 - 2.1 Estimar os valores de ϕ e θ do modelo através de máxima verossimilhança.
 3. Diagnóstico
 - 3.1 Visualizar os resíduos do fitting;
 - 3.2 Plotar histograma, autocorrelação e autocorrelação parcial dos resíduos;
 - 3.3 Testes de estacionaridade e de normalidade.

Assassinato de mulheres

Primeiro, vamos visualizar a série anual. Preecebemos uma clara tendência, com crescimento acentuado ao longo das décadas de 1960 e 1970, bem como um decréscimo após os anos de 1990.



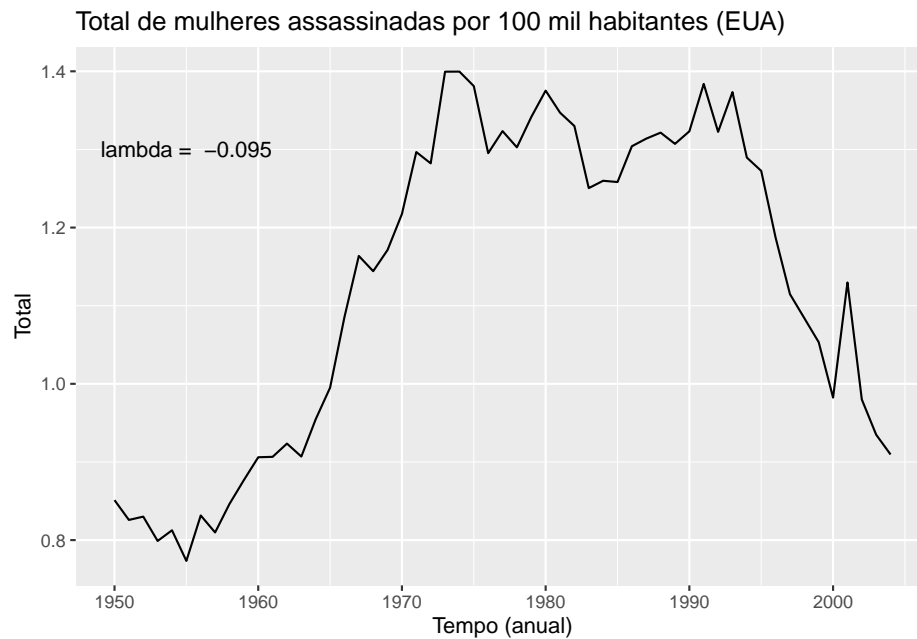
Transformação Box-Cox

Dada a nossa série, é importante que visualizemos a variância ao longo dela. Visualmente ela aparenta não ter variância constante, dado que no início da série, ela aparenta ser menor. Para isso vamos calcular a transformação ótima de Box-Cox.

A transformação de Box-Cox é a seguinte, se $\{y_t\}$ for uma série temporal,

$$y_t^{(\lambda)} = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}, & \text{se } \lambda \neq 0 \\ \log(y_t), & \text{se } \lambda = 0 \end{cases}$$

A escolha de λ ótimo utiliza o método de Guerrero, que escolhe λ que minimize o coeficiente de variação ($c_v = \frac{\sigma}{\mu}$) para subséries de y_t . A seguir podemos conferir o resultado da transformação.



Identificação

O próximo passo do método de Box-Jenkins é identificar os parâmetros p, q, d do modelo ARIMA. O primeiro passo é reduzir para um modelo ARMA estacionário e, para isso, precisamos diferenciar o quanto foi necessário. Vamos utilizar o teste *Augmented Dickey-Fuller* para conferir a estacionariedade.

Teste de Estacionaridade ADF

O teste ADF verifica se há presença de uma raiz unitária na série temporal. A hipótese nula é de que a série possui raiz unitária e é, portanto, não estacionária. Em contrapartida, a alternativa é que a série é estacionária. Vamos considerar $p\text{-valor} < 0.05$ como evidência para rejeitar a hipótese nula, nesse e nos próximos testes que realizarmos.

```
## Warning in adf.test(bcwmurders, alternative = "stationary"): p-value greater
## than printed p-value

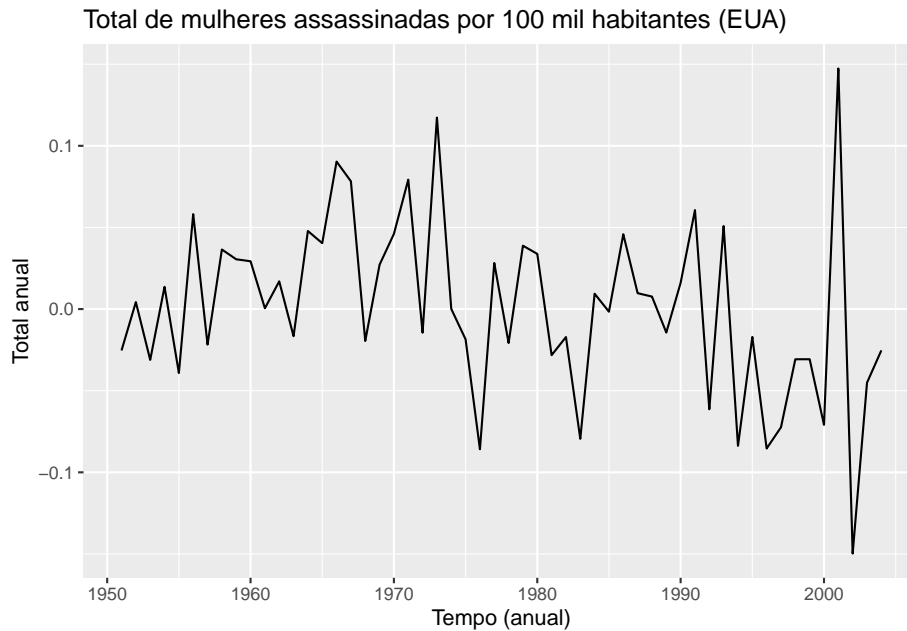
##
## Augmented Dickey-Fuller Test
##
## data: bcwmurders
## Dickey-Fuller = -0.027131, Lag order = 3, p-value = 0.99
## alternative hypothesis: stationary
```

Como $p\text{-valor} \geq 0.05$, não podemos rejeitar a hipótese nula de que a série é não estacionária. Em particular, inferimos que a série é não estacionária, e por isso

vamos diferenciá-la uma vez.

Diferenciando

Tomemos a primeira diferença e, assim, podemos refazer o teste ADF novamente.

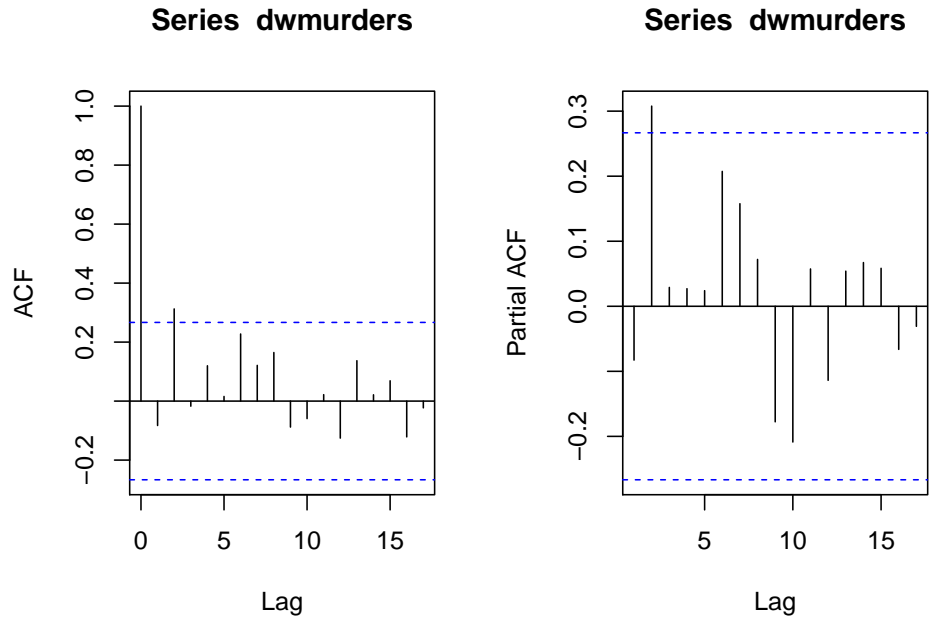


```
##  
## Augmented Dickey-Fuller Test  
##  
## data:  dwmurders  
## Dickey-Fuller = -3.6064, Lag order = 3, p-value = 0.04084  
## alternative hypothesis: stationary
```

Notemos que nosso $p\text{-valor} < 0.05$ e, então, rejeitamos a hipótese nula. Podemos, portanto, seguir a identificação do modelo ARMA.

Autocorrelação e Autocorrelação Parcial

Vamos visualizar o comportamento da ACF e PACF da série diferenciada.



Obtemos um pico no segundo lag, mas não podemos tomar nenhuma conclusão. Poderíamos afirmar, por exemplo, que ambos os gráficos morrem logo, o que indicaria um ruído branco. Mas para não falar isso de forma errônea, vamos considerar os critérios de informação.

Critérios de Informação

Vamos utilizar os critérios de informação estudados:

- Akaike Information Criterion (AIC)
- Akaike Information Criterion with Correction (AICc)
- Bayesian Information Criterion (BIC)

| p | q | AIC | BIC | AICC |
|---|---|-----------|-----------|-----------|
| 0 | 0 | -157.8102 | -153.8322 | -157.5749 |
| 0 | 1 | -156.0378 | -150.0709 | -155.5578 |
| 0 | 2 | -159.8467 | -151.8907 | -159.0303 |
| 0 | 3 | -158.0477 | -148.1028 | -156.7977 |
| 1 | 0 | -156.1778 | -150.2109 | -155.6978 |
| 1 | 1 | -156.6976 | -148.7416 | -155.8813 |
| 1 | 2 | -157.9735 | -148.0285 | -156.7235 |
| 1 | 3 | -156.1456 | -144.2117 | -154.3583 |
| 2 | 0 | -159.4687 | -151.5127 | -158.6523 |
| 2 | 1 | -157.6280 | -147.6831 | -156.3780 |
| 2 | 2 | -158.9166 | -146.9827 | -157.1294 |
| 2 | 3 | -157.2561 | -143.3332 | -154.8213 |

| p | q | AIC | BIC | AICC |
|---|---|-----------|-----------|-----------|
| 3 | 0 | -157.5952 | -147.6503 | -156.3452 |
| 3 | 1 | -155.4728 | -143.5389 | -153.6856 |
| 3 | 2 | -157.4078 | -143.4849 | -154.9730 |
| 3 | 3 | -155.9103 | -139.9984 | -152.7103 |

A partir da tabela, podemos checar alguns pontos: AIC e AICC são bem similares, então vamos ter uma escolha de modelo que respeite ambas as informações. Baseado nelas, o modelo ARMA(0,2) é o que minimiza as informações. Já o BIC é um pouco diferente e em particular ele seleciona o modelo ARMA(0,0) como já havíamos imaginado dados os gráficos. Porém vamos seguir com o modelo ARMA(0,2) dada a liberdade dos coeficientes que ganhamos. Podemos então estimar os coeficientes do modelo ARIMA(0,1,2).

Estimação

Para fazer a estimação, vamos utilizar a função `forecast::Arima` que permite fazer o processo de transformação que fizemos inicialmente com o mesmo método, faz a diferenciação e depois estima os coeficientes. Vejamos o sumário do modelo fittado.

```
## Series: wmurders
## ARIMA(0,1,2)
## Box Cox transformation: lambda= -0.09529835
##
## Coefficients:
##          ma1      ma2
##      -0.0907  0.3770
## s.e.   0.1283  0.1723
##
## sigma^2 estimated as 0.002701: log likelihood=83.92
## AIC=-161.85  AICc=-161.37  BIC=-155.88
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.001070576 0.1996771 0.154647 -0.1341953 4.430353 0.9510067
##              ACF1
## Training set 0.03120906
```

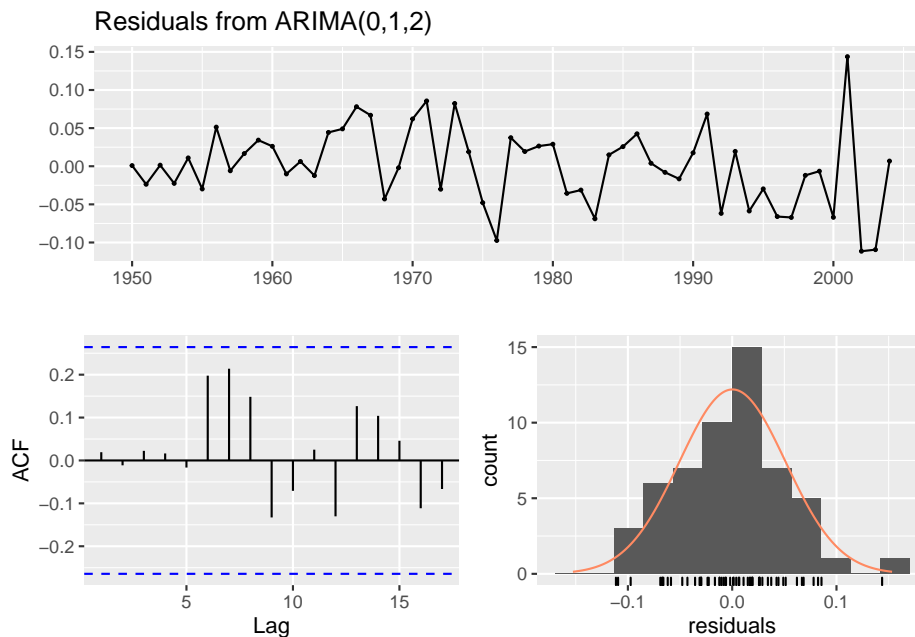
agora temos a estimação do modelo e os correspondentes erro de treino. A partir desse modelo treinado, podemos fazer a diagnose dos resíduos.

Diagnóstico

Vamos utilizar a função `checkresiduals` para fazer a análise inicial dos resíduos gráfica e também utilizaremos os testes Ljung-Box e Jarque Bera.

O teste Ljung-Box testa se as k -correlações do modelo são nulas. Isto é, a hipótese nula é que os dados são distribuídos de forma decorrelacionada, enquanto a alternativa é que alguma autocorrelação é não nula.

Já o teste Jarque-Beta é sobre a assimetria e a curtose da distribuição normal. A hipótese nula é que a assimetria é 0 e excesso de curtose 3.



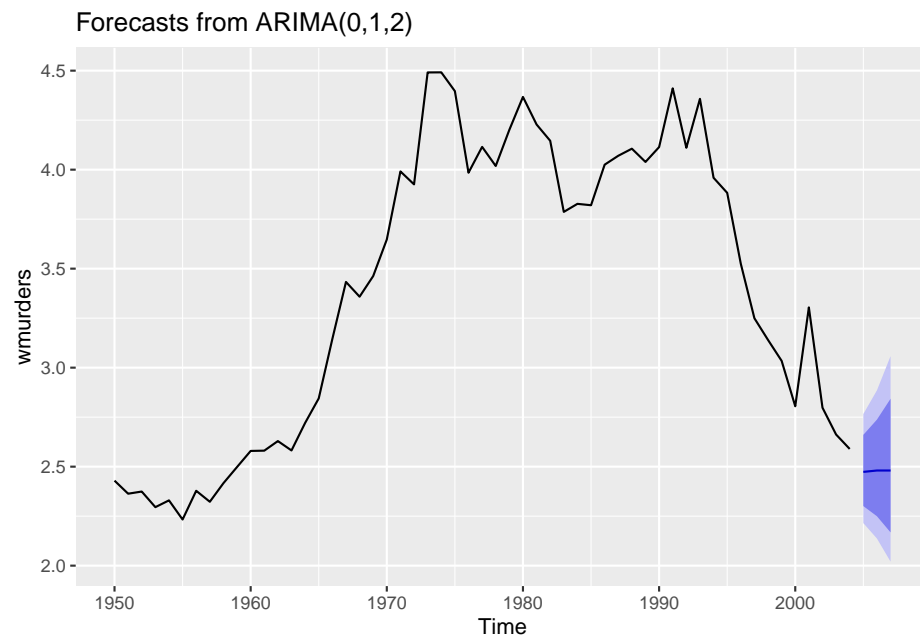
```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,2)
## Q* = 8.6023, df = 8, p-value = 0.3769
##
## Model df: 2.    Total lags used: 10
##
##  Jarque Bera Test
##
## data:  model$residuals
## X-squared = 0.13536, df = 2, p-value = 0.9346
```

A ACF dos resíduos é muito interessante, dados que todas estão entre as faixas de confiança de zero. Já é um indicativo de um processo estacionário, ruído branco. O histograma parece um pouco não simétrico, mas ainda sim a sua suavização tem uma cara de normal bem acentuada. Os resultados dos testes também são interessantes, dado que ambos não rejeitam as suas hipóteses nulas. Ou seja, não temos evidência para rejeitá-las.

Assim estamos contentes com o modelo, e podemos fazer projeções.

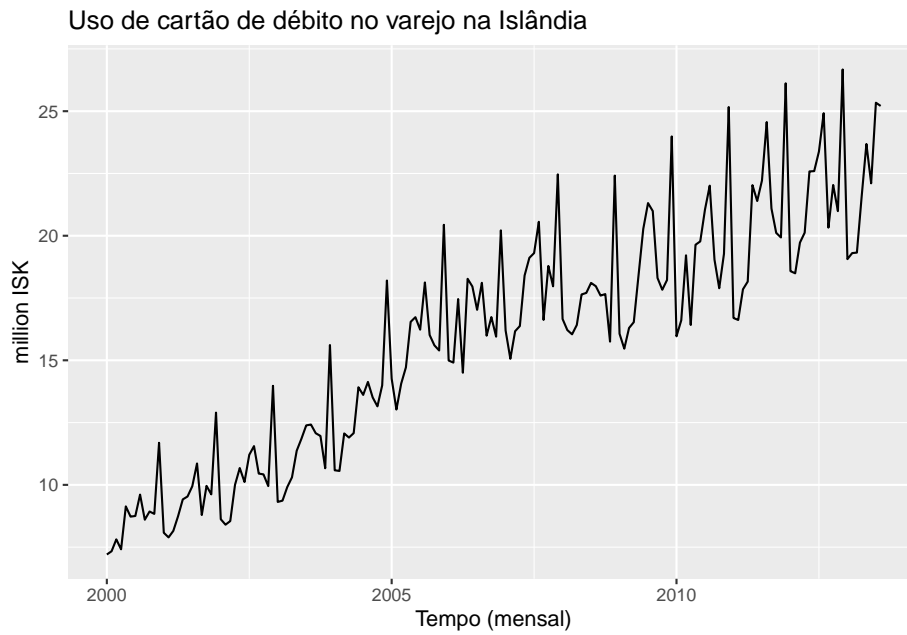
Projeção

Esse é o gráfico da projeção 3 passos a frente.



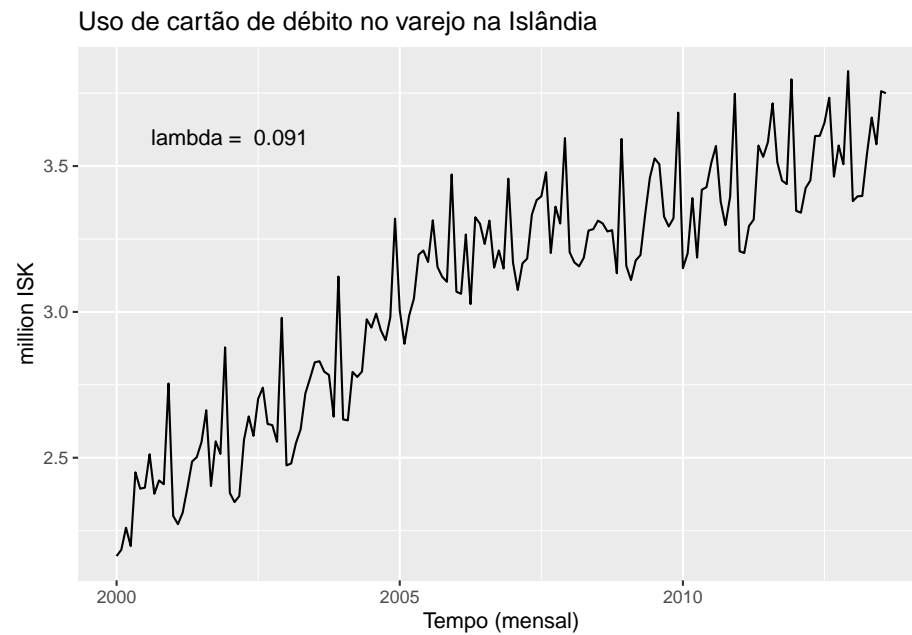
Uso de cartão de débito

Primeiro, vamos visualizar a série. Percebemos uma sutil sazonalidade, aparentemente anual, além de uma clara tendência.



Transformação Box-Cox

Pelo gráfico, parece que teremos que fazer alguma transformação de estabilidade da variância. Para isso, vamos utilizar a transformação Box-Cox, como explicada na série anterior.



Identificação

Agora, com a variância da série estabilizada, podemos fazer o teste de estacionariedade. Observe que o teste ADF possui possibilidade de tendência. Assim

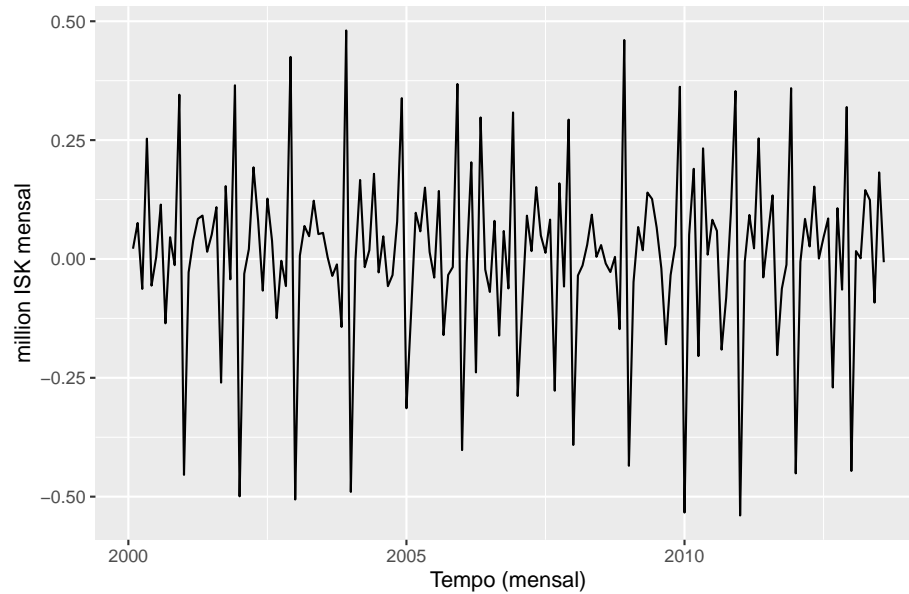
Teste de Estacionaridade ADF

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: bcdebitcards  
## Dickey-Fuller = -3.0873, Lag order = 5, p-value = 0.1227  
## alternative hypothesis: stationary
```

Como $p\text{-valor} \geq 0.05$, não podemos rejeitar a hipótese nula de que a série é não estacionária. Portanto, vamos usar a primeira diferenciação.

Diferenciando

Uso de cartão de débito no varejo na Islândia



```
## Warning in adf.test(ddebitcards): p-value smaller than printed p-value
```

```
##
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

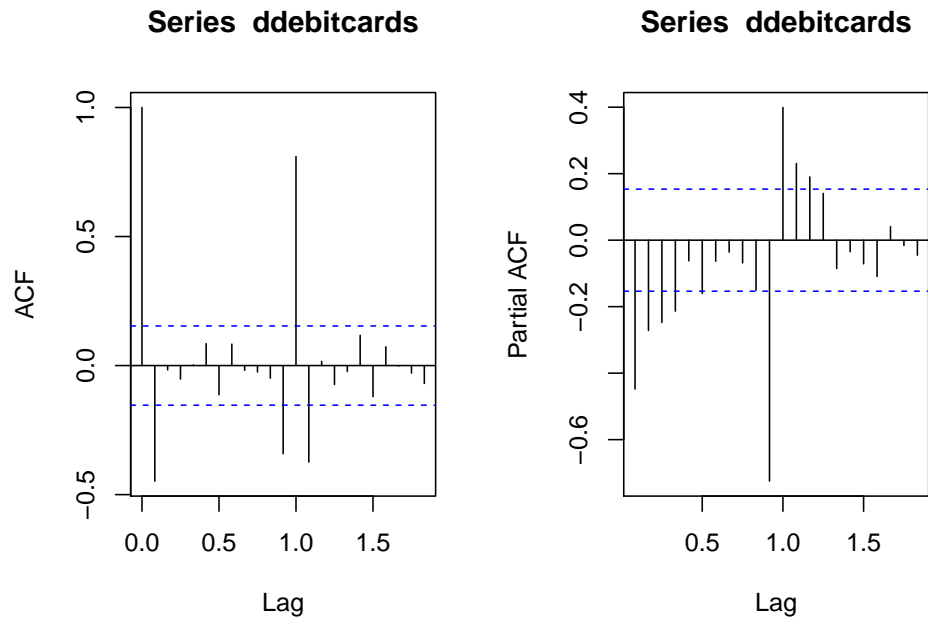
```
## data: ddebitcards
```

```
## Dickey-Fuller = -8.3838, Lag order = 5, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

Assim, podemos rejeitar a hipótese nula, o que suporta a ideia de que a série é estacionária. Vamos considerar os gráficos da ACF e PACF.

ACF e PACF



Percebemos dois fatores bem destacados: uma grande correlação quando o $\text{Lag} = 12$, o que indica que existe uma sazonalidade anual; e que a PACF decresce exponencialmente, enquanto a ACF morre após $\text{Lag} = 1$, o que nos levaria a um modelo $\text{MA}(1)$.

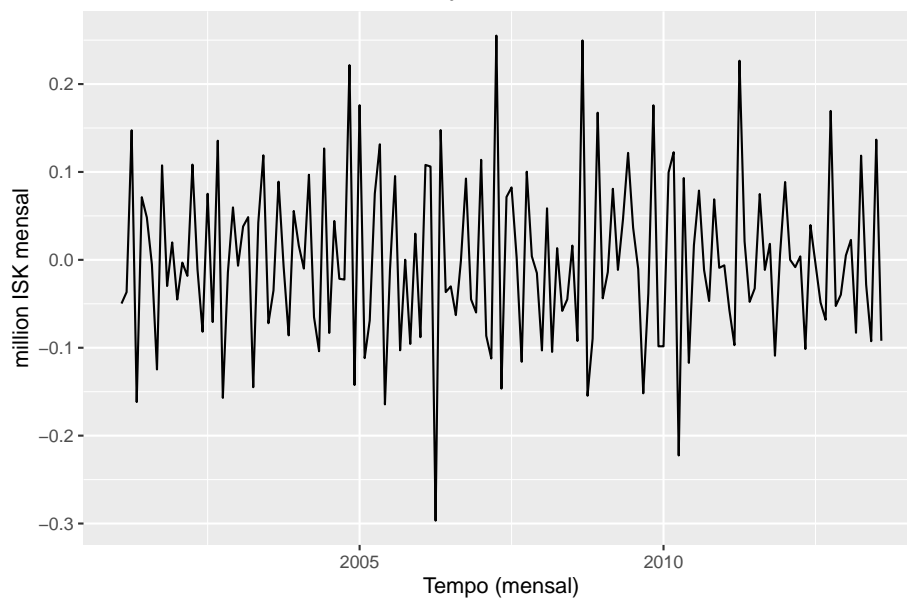
Antes disso, vamos verificar a sazonalidade anual, como é nossa hipótese. Para isso, vamos utilizar o teste Kruskal-Wallis, como já utilizado em outros trabalhos.

```
##
## Kruskal-Wallis rank sum test
##
## data: ddebitcards and g
## Kruskal-Wallis chi-squared = 113.17, df = 11, p-value < 2.2e-16
```

Como o $p\text{-valor} < 0.05$, rejeitamos a hipótese nula de que os coeficientes $\alpha_1 = \dots = \alpha_{11} = 0$ e, portanto, existe sazonalidade determinística. Para removê-la, precisamos usar uma diferenciação anual. Segue o resultado:

Diferenciação Sazonal

Uso de cartão de débito no varejo na Islândia



```
## Warning in adf.test(ddebitcards): p-value smaller than printed p-value
```

```
##
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

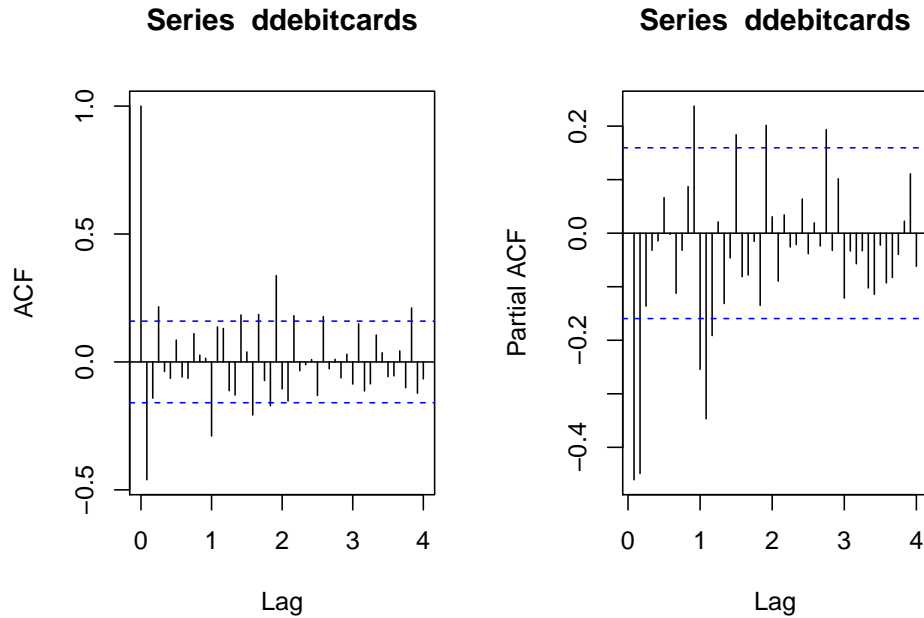
```
## data: ddebitcards
```

```
## Dickey-Fuller = -5.4939, Lag order = 5, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

O teste ADF confirma que ainda temos uma série estacionária.

Observe que ainda temos uma série estacionária. Vejamos a ACF e PACF novamente:



Os picos dos lags 12 e 24 diminuíram consideravelmente, entretanto eles ainda existem e mesmo diferenciando mais uma ou duas vezes, não foi possível resolvê-lo. Isso pode ser um indicativo de que haja algum termo autorregressivo e de média móvel sazonal, os modelos SARIMA que ainda não foram estudados em aula. Para isso, vou fazer duas considerações:

1. Um modelo sem termos sazonais AR e MA, apenas uma diferença sazonal;
2. Um modelo com um termo MA.

Vamos compará-los usando os critérios de informação.

Observação: Não foi possível adicionar termo AR sazonal em alguns modelos dado problema de estacionariedade. Como não nos aprofundamos desse tema, descartei da análise.

Critérios de Informação

| p | q | AIC | BIC | AICc | AIC (S) | BIC (S) | AICc (S) |
|---|---|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0 | -277.1636 | -271.1291 | -277.0825 | -319.8353 | -310.7835 | -319.6720 |
| 0 | 1 | -339.0449 | -329.9930 | -338.8816 | -388.0190 | -375.9499 | -387.7451 |
| 0 | 2 | -339.0065 | -326.9374 | -338.7326 | -389.5038 | -374.4174 | -389.0900 |
| 0 | 3 | -342.2090 | -327.1226 | -341.7952 | -388.8129 | -370.7092 | -388.2296 |
| 0 | 4 | -342.0094 | -323.9057 | -341.4261 | -392.6817 | -371.5608 | -391.8985 |
| 0 | 5 | -345.7949 | -324.6739 | -345.0116 | -391.0359 | -366.8977 | -390.0218 |
| 1 | 0 | -311.2501 | -302.1983 | -311.0869 | -361.7020 | -349.6329 | -361.4280 |
| 1 | 1 | -337.9786 | -325.9095 | -337.7046 | -388.1459 | -373.0595 | -387.7321 |
| 1 | 2 | -338.9584 | -323.8720 | -338.5446 | -387.8668 | -369.7632 | -387.2835 |

| p | q | AIC | BIC | AICc | AIC (S) | BIC (S) | AICc (S) |
|---|---|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 3 | -341.1438 | -323.0401 | -340.5604 | -389.5604 | -368.4395 | -388.7772 |
| 1 | 4 | -342.3731 | -321.2522 | -341.5899 | -391.0381 | -366.8999 | -390.0241 |
| 1 | 5 | -343.8243 | -319.6860 | -342.8102 | -389.0122 | -361.8567 | -387.7356 |
| 2 | 0 | -343.2455 | -331.1764 | -342.9716 | -394.4030 | -379.3166 | -393.9892 |
| 2 | 1 | -343.9712 | -328.8848 | -343.5574 | -393.5461 | -375.4424 | -392.9628 |
| 2 | 2 | -342.0010 | -323.8973 | -341.4177 | -391.7810 | -370.6600 | -390.9978 |
| 2 | 3 | -340.0692 | -318.9482 | -339.2859 | -406.4475 | -382.3093 | -405.4334 |
| 2 | 4 | -345.1572 | -321.0189 | -344.1431 | -389.0559 | -361.9004 | -387.7793 |
| 2 | 5 | -354.2751 | -327.1196 | -352.9985 | -387.2402 | -357.0674 | -385.6688 |
| 3 | 0 | -343.8279 | -328.7415 | -343.4141 | -393.2912 | -375.1875 | -392.7078 |
| 3 | 1 | -341.9907 | -323.8870 | -341.4073 | -391.9992 | -370.8782 | -391.2160 |
| 3 | 2 | -343.4891 | -322.3682 | -342.7059 | -390.0152 | -365.8769 | -389.0011 |
| 3 | 3 | -341.8366 | -317.6984 | -340.8226 | -405.2228 | -378.0673 | -403.9462 |
| 3 | 4 | -339.8344 | -312.6788 | -338.5578 | -402.6506 | -372.4778 | -401.0792 |
| 3 | 5 | -350.4613 | -320.2885 | -348.8899 | -401.9700 | -368.7799 | -400.0707 |
| 4 | 0 | -341.9974 | -323.8937 | -341.4140 | -391.7803 | -370.6594 | -390.9971 |
| 4 | 1 | -340.0049 | -318.8839 | -339.2217 | -389.8239 | -365.6857 | -388.8099 |
| 4 | 2 | -341.7904 | -317.6522 | -340.7763 | -394.6122 | -367.4567 | -393.3356 |
| 4 | 3 | -339.8401 | -312.6846 | -338.5635 | -403.7273 | -373.5545 | -402.1559 |
| 4 | 4 | -350.3646 | -320.1918 | -348.7932 | -401.7038 | -368.5137 | -399.8045 |
| 4 | 5 | -353.3164 | -320.1263 | -351.4171 | -398.7476 | -362.5402 | -396.4867 |
| 5 | 0 | -340.0364 | -318.9154 | -339.2531 | -389.9467 | -365.8084 | -388.9326 |
| 5 | 1 | -338.0020 | -313.8638 | -336.9879 | -388.0515 | -360.8960 | -386.7749 |
| 5 | 2 | -341.2724 | -314.1169 | -339.9958 | -388.9848 | -358.8121 | -387.4134 |
| 5 | 3 | -344.0011 | -313.8283 | -342.4297 | -385.5135 | -352.3234 | -383.6142 |
| 5 | 4 | -345.4699 | -312.2798 | -343.5706 | -393.8582 | -357.6509 | -391.5974 |
| 5 | 5 | -349.3232 | -313.1158 | -347.0623 | -403.4238 | -364.1992 | -400.7669 |

Baseado no AIC e AICc, o modelo é ARMA(2,3)(0,1) é o escolhido. Felizmente o BIC concorda com essa hipótese, então esse é o modelo escolhido.

Estimação

Estimamos os parâmetros do modelo, considerando uma diferenciação sazonal.

```
## Series: debitcards
## ARIMA(2,1,3)(0,1,1)[12]
## Box Cox transformation: lambda= 0.09078485
##
## Coefficients:
##          ar1          ar2          ma1          ma2          ma3          sma1
##        -1.1586    -0.9996    0.5464    0.2722    -0.6206    -0.8575
## s.e.      0.0057      0.0012    0.1328    0.1434      0.0669      0.1183
##
```

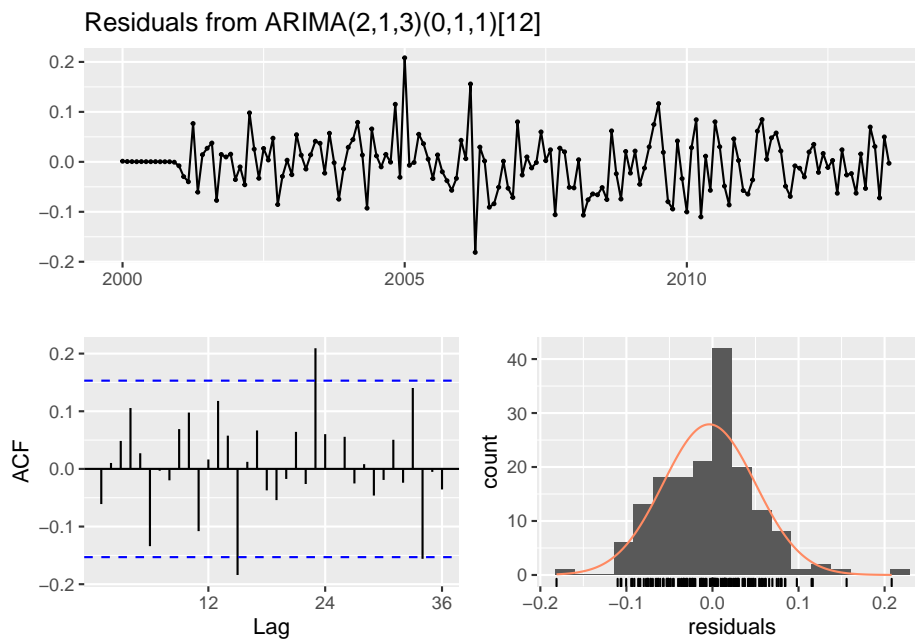
```
## sigma^2 estimated as 0.003259: log likelihood=210.31
## AIC=-406.61 AICc=-405.83 BIC=-385.49
##
## Training set error measures:
##           ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.05699844 0.687529 0.5100688 -0.3707241 3.139493 0.3980881
##           ACF1
## Training set -0.02221451
```

Temos, portanto, o resultado da estimação.

Diagnóstico

Aqui podemos conferir os resíduos. É importante destacar a aparência normal dos resíduos, apesar de que existe uma aparente assimetria. Porém a ACF ainda apresenta picos, o que é um pouco complicado, mas compreensível.

O test Ljung-Box corrobora esse fato, rejeitando a hipótese nula de que os coeficientes de correlação são iguais. Já Jarque Bera também rejeita a hipótese nula, o que nos dá evidência suficiente para rejeitar a normalidade dos resíduos, dada a aparente assimetria.

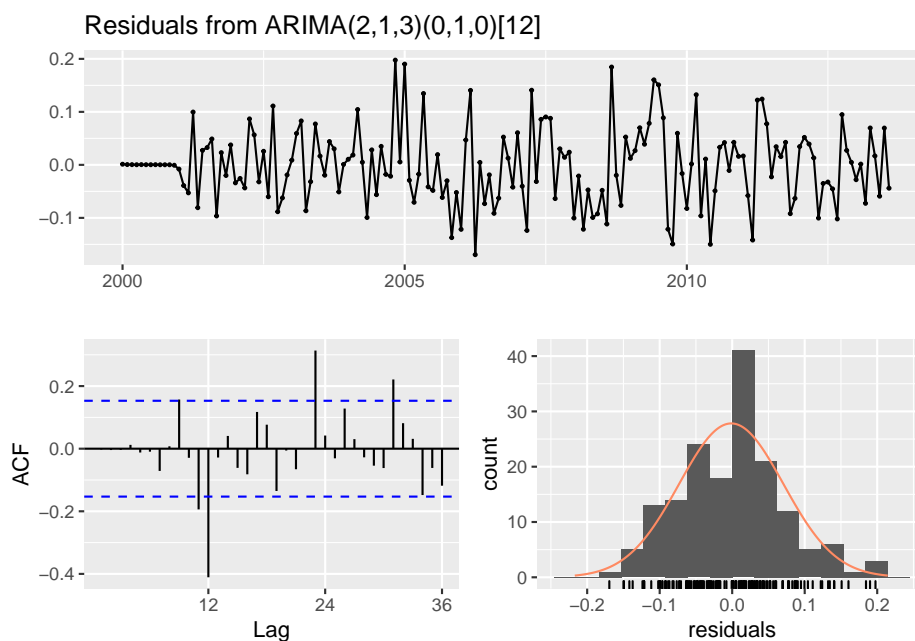


```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(2,1,3)(0,1,1)[12]
## Q* = 31.957, df = 18, p-value = 0.02225
```



```
##
## Model df: 6.    Total lags used: 24
##
## Jarque Bera Test
##
## data:  model$residuals
## X-squared = 14.644, df = 2, p-value = 0.000661
```

Apenas olhando os resíduos, fica muito difícil de propor alguma mudança. Todavia uma proposição interessante seria retirar a média móvel sazonal.



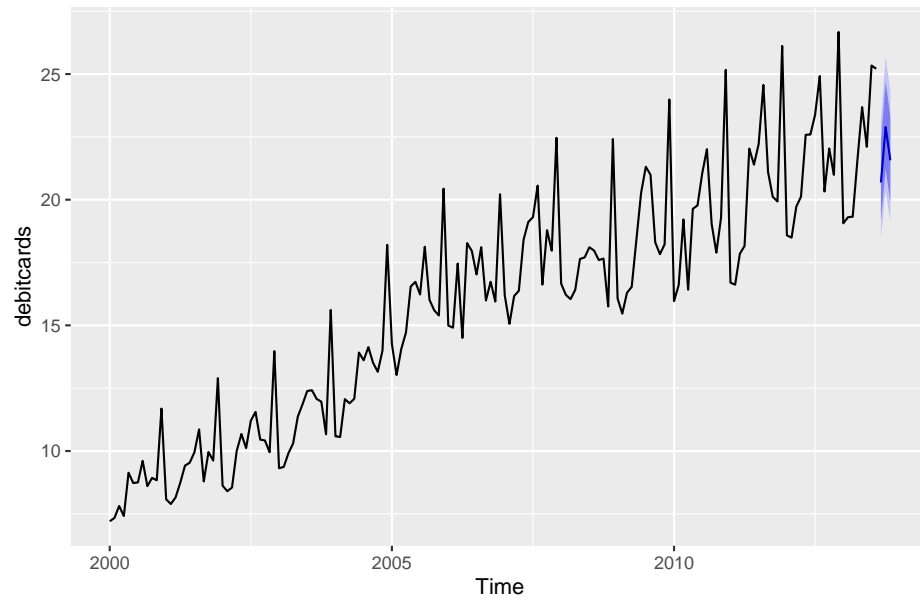
```
##
## Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,3)(0,1,0)[12]
## Q* = 71.938, df = 19, p-value = 4.368e-08
##
## Model df: 5.    Total lags used: 24
##
## Jarque Bera Test
##
## data:  model$residuals
## X-squared = 1.991, df = 2, p-value = 0.3695
```

Apesar de termos piorado o lag 12, que ainda continua com o pico, resolvemos o problema da normalidade, de certa forma.

Projeção

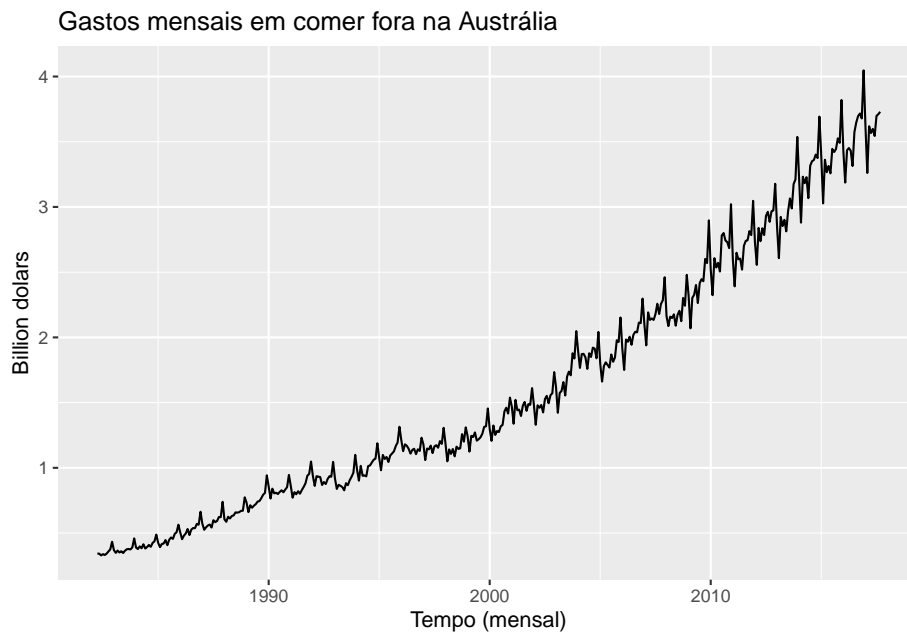
Agora vamos conferir as projeções três passos a frente.

Forecasts from ARIMA(2,1,3)(0,1,0)[12]



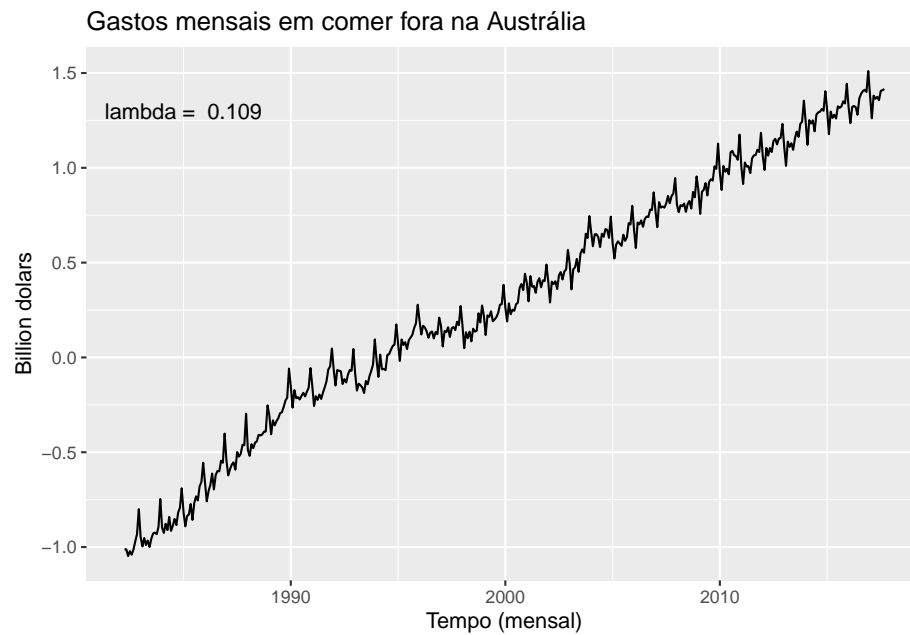
Comer fora na Austrália

Primeiro, vamos visualizar a série. Percebemos que há uma aparente sazonalidade, além de que a variância parece aumentar à medida que se aumenta o gasto. Em geral, a tendência é de crescimento.



Transformação de Box-Cox

É clara a necessidade de uma tentativa de estabilização da variância. Vemos que, aparentemente, a série teve sua variância estabilizada.



Identificação

Com a variância estabilizada, seguimos com um teste ADF.

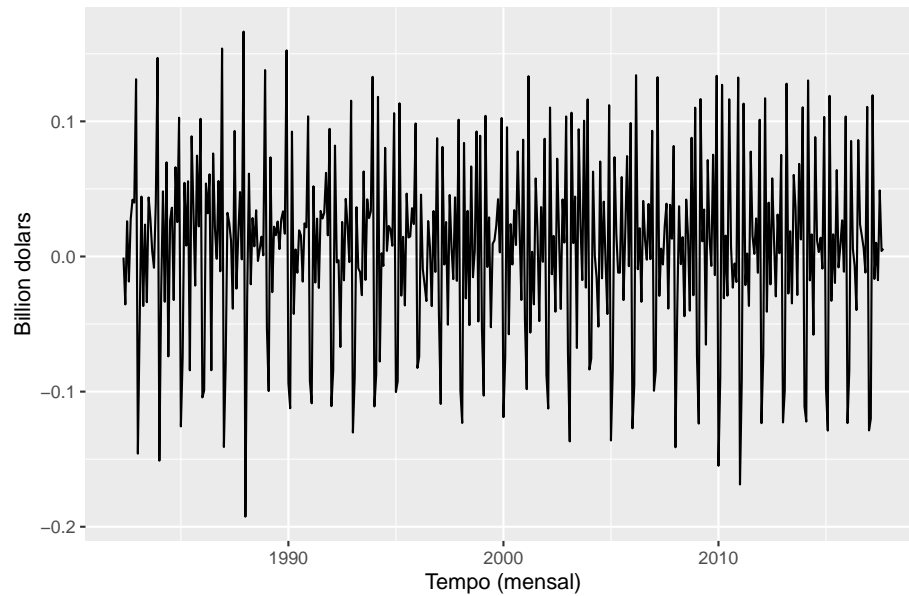
Teste de Estacionaridade ADF

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: bcauscafe  
## Dickey-Fuller = -3.3838, Lag order = 7, p-value = 0.05669  
## alternative hypothesis: stationary
```

Como $p\text{-valor} \geq 0.05$ (por pouco), não rejeitamos a hipótese nula de não estacionariedade. Vamos portanto diferenciar a série.

Diferenciação

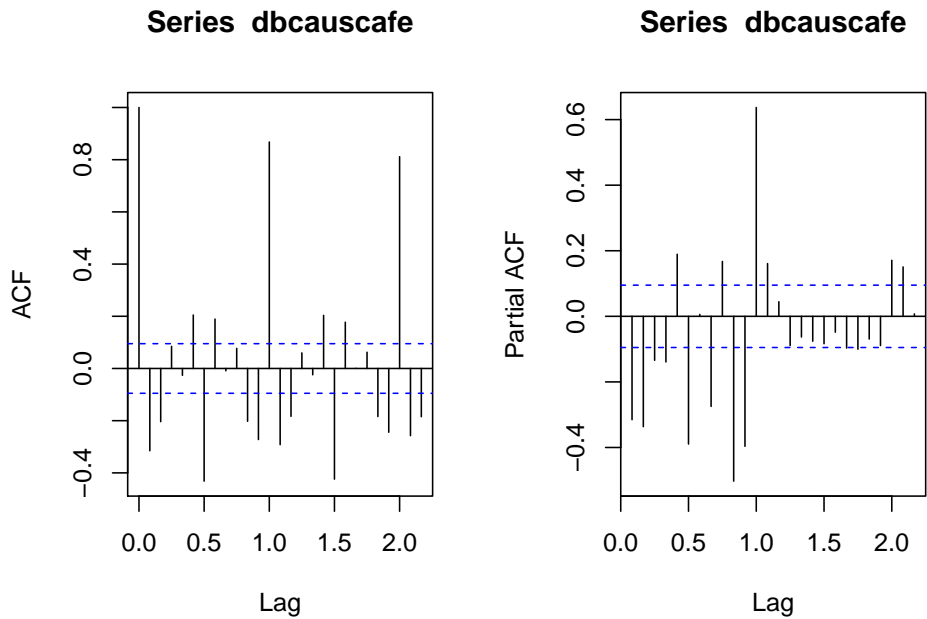
Gastos mensais em comer fora na Austrália



```
## Warning in adf.test(dbcauscafe): p-value smaller than printed p-value
##
## Augmented Dickey-Fuller Test
##
## data: dbcauscafe
## Dickey-Fuller = -13.205, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

Receitamos a hipótese nula de não estacionariedade. Vejamos a ACF e a PACF da série:

ACF e PACF

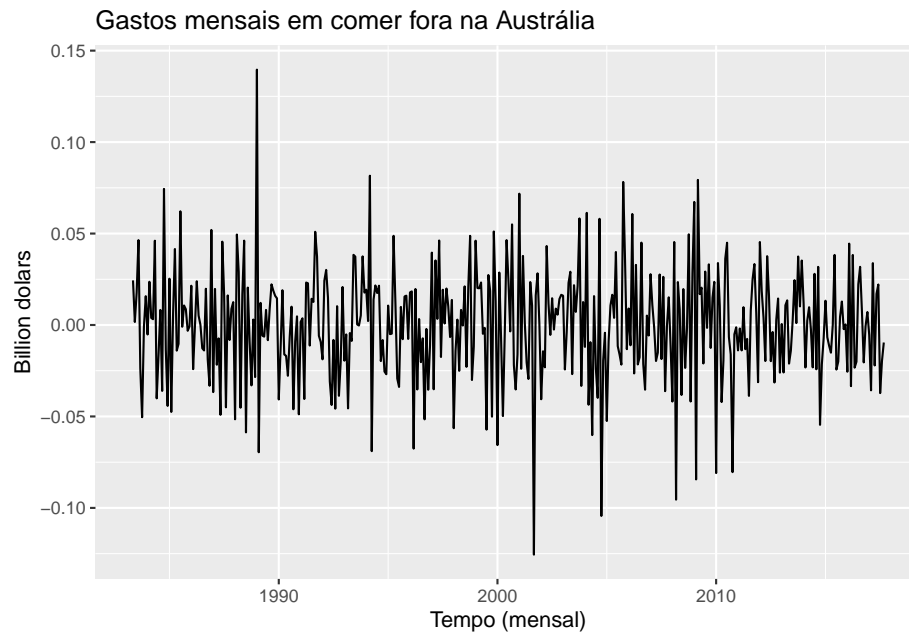


Vemos picos de autocorrelação para lags múltiplos de 6, o indica uma sazonalidade de 6 ou 12 meses. Vamos testar a sazonalidade:

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: dbcausafe and g  
## Kruskal-Wallis chi-squared = 358.12, df = 11, p-value < 2.2e-16
```

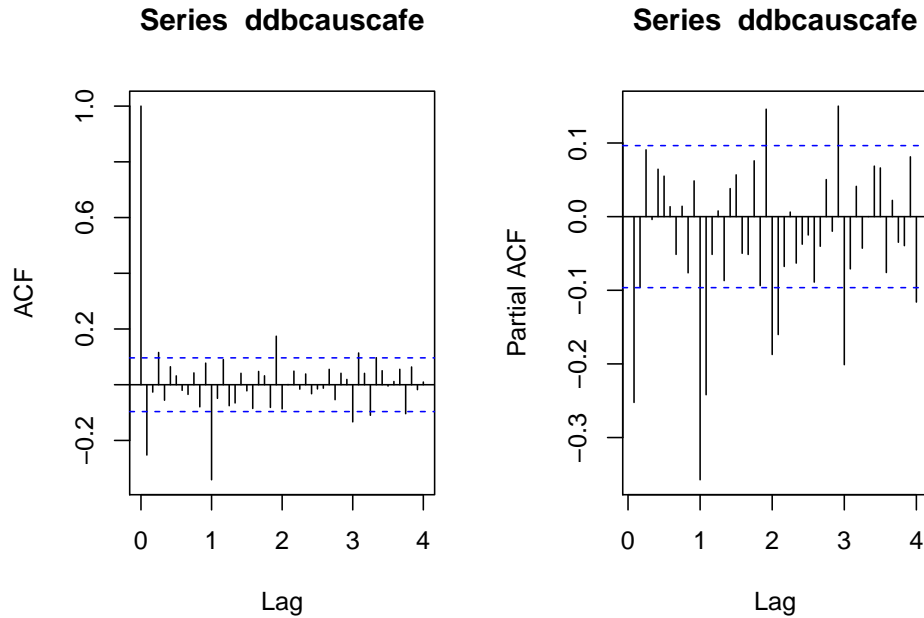
Rejeitamos a não existência de sazonalidade determinística. Dessa forma, vamos realizar uma diferenciação sazonal, de lag 12.

Diferenciação Sazonal



```
## Warning in adf.test(ddbcauscafe): p-value smaller than printed p-value
##
## Augmented Dickey-Fuller Test
##
## data: ddbcauscafe
## Dickey-Fuller = -6.7817, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

Ainda concluímos que a série é estacionária. Vejamos a ACF e PACF novamente:



Os picos diminuíram de forma considerável, tanto na ACF quanto na PACF.

É muito difícil tomar decisões sobre a ordem de um modelo ARMA apenas se baseando nos gráficos da ACF e da PACF. No caso da série com diferenciação sazonal, poderíamos argumentar senóides amortecidas na PACF, sendo portanto um MA(1). Porém, temos dificuldade de aceitar isso.

Dessa forma, vamos tentar um modelo que leva em consideração MA(1) sazonal, por causa dos picos na PACF, e outro que fica apenas na diferenciação sazonal. Os parâmetros do modelo não sazonal serão escolhidos por critério de informação.

Critérios de Informação

| p | q | AIC | BIC | AICc | AIC (S) | BIC (S) | AICc (S) |
|---|---|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0 | -1694.497 | -1686.450 | -1694.467 | -1829.028 | -1816.958 | -1828.969 |
| 0 | 1 | -1721.607 | -1709.537 | -1721.548 | -1882.555 | -1866.461 | -1882.457 |
| 0 | 2 | -1719.999 | -1703.905 | -1719.901 | -1881.330 | -1861.213 | -1881.183 |
| 0 | 3 | -1722.701 | -1702.584 | -1722.553 | -1882.765 | -1858.624 | -1882.558 |
| 0 | 4 | -1720.781 | -1696.640 | -1720.574 | -1882.661 | -1854.497 | -1882.385 |
| 0 | 5 | -1723.952 | -1695.788 | -1723.676 | -1886.088 | -1853.900 | -1885.731 |
| 0 | 6 | -1732.125 | -1699.938 | -1731.769 | -1884.127 | -1847.916 | -1883.681 |
| 1 | 0 | -1719.641 | -1707.570 | -1719.582 | -1875.797 | -1859.703 | -1875.699 |
| 1 | 1 | -1719.815 | -1703.722 | -1719.717 | -1881.042 | -1860.925 | -1880.895 |
| 1 | 2 | -1720.686 | -1700.569 | -1720.539 | -1880.605 | -1856.464 | -1880.398 |
| 1 | 3 | -1733.989 | -1709.848 | -1733.782 | -1884.840 | -1856.676 | -1884.563 |
| 1 | 4 | -1732.899 | -1704.735 | -1732.623 | -1883.206 | -1851.018 | -1882.850 |

| p | q | AIC | BIC | AICc | AIC (S) | BIC (S) | AICc (S) |
|---|---|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 5 | -1733.973 | -1701.785 | -1733.616 | -1884.111 | -1847.900 | -1883.665 |
| 1 | 6 | -1734.654 | -1698.443 | -1734.207 | -1887.712 | -1847.477 | -1887.164 |
| 2 | 0 | -1721.492 | -1705.398 | -1721.394 | -1884.914 | -1864.796 | -1884.766 |
| 2 | 1 | -1734.655 | -1714.538 | -1734.507 | -1887.723 | -1863.583 | -1887.517 |
| 2 | 2 | -1733.161 | -1709.021 | -1732.954 | -1885.750 | -1857.586 | -1885.474 |
| 2 | 3 | -1734.279 | -1706.115 | -1734.003 | -1883.751 | -1851.563 | -1883.394 |
| 2 | 4 | -1732.508 | -1700.321 | -1732.152 | -1884.985 | -1848.774 | -1884.539 |
| 2 | 5 | -1732.029 | -1695.818 | -1731.583 | -1906.637 | -1866.402 | -1906.090 |
| 2 | 6 | -1778.275 | -1738.041 | -1777.728 | -1885.763 | -1841.505 | -1885.104 |
| 3 | 0 | -1722.835 | -1702.718 | -1722.688 | -1885.999 | -1861.858 | -1885.792 |
| 3 | 1 | -1721.183 | -1697.042 | -1720.976 | -1885.751 | -1857.587 | -1885.475 |
| 3 | 2 | -1719.813 | -1691.649 | -1719.537 | -1883.724 | -1851.536 | -1883.367 |
| 3 | 3 | -1753.880 | -1721.693 | -1753.524 | -1908.421 | -1872.210 | -1907.975 |
| 3 | 4 | -1761.487 | -1725.276 | -1761.040 | -1909.598 | -1869.364 | -1909.051 |
| 3 | 5 | -1763.065 | -1722.831 | -1762.518 | -1908.246 | -1863.988 | -1907.588 |
| 3 | 6 | -1762.062 | -1717.804 | -1761.404 | -1906.712 | -1858.431 | -1905.932 |
| 4 | 0 | -1720.843 | -1696.703 | -1720.637 | -1885.585 | -1857.421 | -1885.309 |
| 4 | 1 | -1732.162 | -1703.998 | -1731.886 | -1883.766 | -1851.579 | -1883.410 |
| 4 | 2 | -1742.136 | -1709.948 | -1741.779 | -1908.675 | -1872.464 | -1908.228 |
| 4 | 3 | -1757.348 | -1721.137 | -1756.902 | -1887.841 | -1847.607 | -1887.294 |
| 4 | 4 | -1766.735 | -1726.500 | -1766.187 | -1908.206 | -1863.948 | -1907.548 |
| 4 | 5 | -1752.771 | -1708.513 | -1752.113 | -1907.964 | -1859.682 | -1907.184 |
| 4 | 6 | -1755.463 | -1707.182 | -1754.683 | -1906.577 | -1854.272 | -1905.665 |
| 5 | 0 | -1720.577 | -1692.413 | -1720.301 | -1884.096 | -1851.909 | -1883.740 |
| 5 | 1 | -1719.387 | -1687.199 | -1719.030 | -1882.885 | -1846.674 | -1882.439 |
| 5 | 2 | -1741.081 | -1704.870 | -1740.635 | -1881.896 | -1841.662 | -1881.349 |
| 5 | 3 | -1741.762 | -1701.528 | -1741.215 | -1886.648 | -1842.391 | -1885.990 |
| 5 | 4 | -1742.233 | -1697.975 | -1741.574 | -1910.147 | -1861.866 | -1909.367 |
| 5 | 5 | -1781.319 | -1733.038 | -1780.539 | -1883.090 | -1830.785 | -1882.177 |
| 5 | 6 | -1783.148 | -1730.843 | -1782.236 | -1907.357 | -1851.029 | -1906.302 |
| 6 | 0 | -1719.911 | -1687.724 | -1719.555 | -1883.844 | -1847.633 | -1883.398 |
| 6 | 1 | -1717.941 | -1681.730 | -1717.494 | -1881.918 | -1841.683 | -1881.370 |
| 6 | 2 | -1741.257 | -1701.022 | -1740.710 | -1905.279 | -1861.021 | -1904.621 |
| 6 | 3 | -1741.352 | -1697.094 | -1740.694 | -1877.994 | -1829.713 | -1877.214 |
| 6 | 4 | -1765.679 | -1717.397 | -1764.899 | -1901.396 | -1849.091 | -1900.483 |
| 6 | 5 | -1782.882 | -1730.577 | -1781.969 | -1906.442 | -1850.113 | -1905.386 |
| 6 | 6 | -1780.383 | -1724.055 | -1779.328 | -1903.056 | -1842.704 | -1901.847 |

Baseando-se no AIC e no AICc, verificamos que um modelo ARMA(5, 4)(0, 1) é o escolhido. É válido citar que seu BIC não é o menor, porém temos de tomar decisões.

Estimação

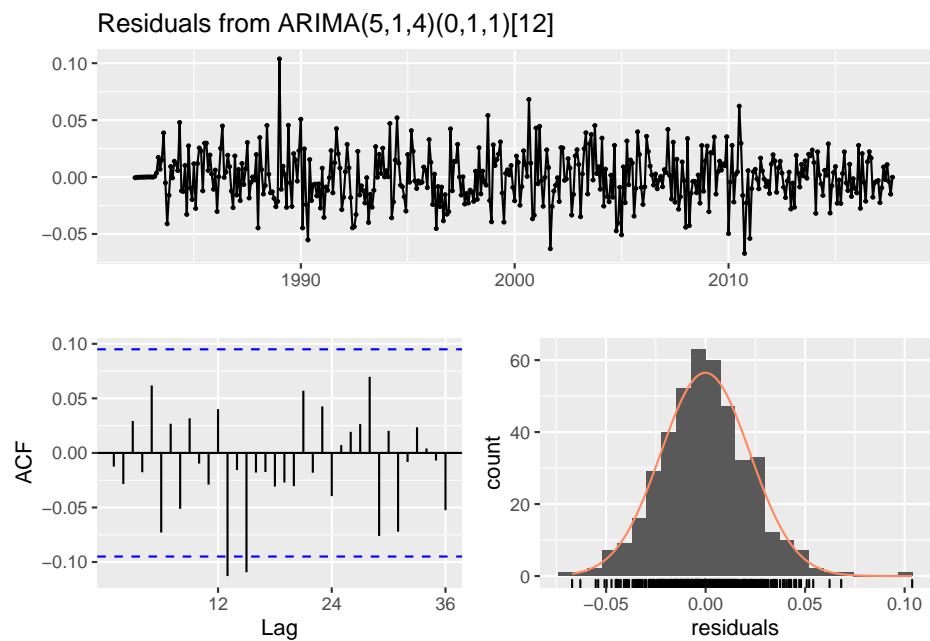
Vamos estimar um modelo ARIMA(5,1,4)(0,1,1)[12].

```
## Series: auscafe
## ARIMA(5,1,4)(0,1,1)[12]
## Box Cox transformation: lambda= 0.109056
##
## Coefficients:
##          ar1          ar2          ar3          ar4          ar5          ma1          ma2          ma3
##          0.0069 -0.2043  0.2668 -0.8504 -0.2645 -0.3420  0.2563 -0.3357
## s.e.    0.0491  0.0282  0.0292  0.0269  0.0494  0.0201  0.0121  0.0160
##          ma4          sma1
##          0.9957 -0.8124
## s.e.    0.0220  0.0330
##
## sigma^2 estimated as 0.0005269:  log likelihood=967.04
## AIC=-1912.07  AICc=-1911.42  BIC=-1867.82
##
## Training set error measures:
##              ME          RMSE          MAE          MPE          MAPE          MASE
## Training set -0.0006603977  0.03505846  0.02513782 -0.02891303  1.67619  0.2405844
##              ACF1
## Training set -0.02669772
```

Esse é o resultado da estimação.

Diagnóstico

Os resíduos não aparentam se diferenciar do que seriam resíduos normais. Sua distribuição fica próxima de uma distribuição normal. A ACF parece razoavelmente comportada. Porém, vemos que rejeitamos a normalidade, mas não rejeitamos a não correlação.



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(5,1,4)(0,1,1)[12]
## Q* = 23.536, df = 14, p-value = 0.05208
##
## Model df: 10.    Total lags used: 24
##
##  Jarque Bera Test
##
## data:  model$residuals
## X-squared = 19.83, df = 2, p-value = 4.943e-05
```

Decidimos por escolher esse modelo.

Projeção

Vamos conferir as projeções três passos a frente.

Forecasts from ARIMA(5,1,4)(0,1,1)[12]

