

# Séries Temporais em $\mathbf{R}^*$

Ricardo Lima<sup>†</sup>

---

\*Este arquivo ainda encontra-se em fase de construção. Logo, não deve ser citado e nem usado como referência para trabalhos acadêmicos.

<sup>†</sup>Economista pela Universidade Federal Fluminense e Mestrando em economia pela Universidade de Estocolmo. Email para contato: [ricardo.lima@ne.su.se](mailto:ricardo.lima@ne.su.se).

# Sumário

<b>1</b>	<b>Introdução</b>	<b>5</b>
1.1	Por que usar o <b>R</b> ?	5
1.2	O que é uma série de tempo?	7
1.3	Porque analisar séries de tempo?	7
<b>2</b>	<b>Começando os trabalhos</b>	<b>8</b>
2.1	Importando e visualizando os dados	8
<b>3</b>	<b>Fatos estilizados do Ciclo de Negócios</b>	<b>11</b>
3.1	Removendo tendências de séries univariadas	11
3.1.1	Removendo sazonalidade	11
3.1.2	Tendência de longo prazo	16
3.2	Calculando correlações cíclicas	22
3.3	Defasagens no <b>R</b>	23
3.4	Calculando a volatilidade cíclica	24
<b>4</b>	<b>Modelos univariados</b>	<b>26</b>
4.1	Alguns conceitos básicos	27
4.2	Processos de médias móveis	29
4.3	Processos autoregressivos	31
4.3.1	O operador de defasagem, $L$	33
4.3.2	Passeio aleatório	34
4.3.3	Passeio aleatório com <i>drift</i>	35
4.3.4	Processos tendência-estacionários	37
<b>5</b>	<b>Função de autocorrelação</b>	<b>38</b>
<b>6</b>	<b>Função de autocorrelação parcial</b>	<b>40</b>
6.1	Um breve resumo dos resultados	41
<b>7</b>	<b>Processos ARMA</b>	<b>43</b>
7.1	Estimando modelos <i>ARIMA</i> - Intro	43
7.1.1	A metodologia <i>Box-Jenkins</i> - Intro	43

<b>8</b>	<b>Testes de estacionaridade</b>	<b>45</b>
8.1	O teste de <i>Dickey-Fuller</i> . . . . .	45
8.2	O Poder do teste de <i>Dickey-Fuller</i> de raíz unitária . . . . .	51
8.3	O teste de <i>Dickey-Fuller</i> Aumentado . . . . .	52
8.4	O teste de <i>Phillips-Perron</i> . . . . .	53
8.5	O teste <i>ADF</i> sequencial . . . . .	54
<b>9</b>	<b>Estimando modelos ARIMA</b>	<b>64</b>
9.1	A metodologia <i>Box-Jenkins</i> . . . . .	64
9.2	A metodologia <i>Box-Jenkins</i> - Um exemplo prático . . . . .	65
9.2.1	Ordem de integração . . . . .	66
9.2.2	Escolhendo $p$ e $q$ . . . . .	68
9.2.3	Estimando um modelo ARMA . . . . .	69
9.2.4	Diagnósticos dos resíduos . . . . .	71
<b>10</b>	<b>Previsão com modelos ARIMA</b>	<b>77</b>
10.1	Usando modelos <i>ARIMA</i> para construir previsões . . . . .	77
10.1.1	O caso do processo autoregressivo . . . . .	77
10.1.2	O caso do processo de médias móveis . . . . .	78
10.2	Prevendo com o <b>R</b> . . . . .	79
10.3	Avaliando previsões . . . . .	81
<b>11</b>	<b>Modelos multivariados</b>	<b>86</b>
11.1	Modelos dinâmicos com variáveis estacionárias . . . . .	86
11.2	Modelos de correção de erro . . . . .	88
11.3	Estimando e testando . . . . .	88
11.4	Causalidade de Granger . . . . .	88
11.5	Propriedades dos estimadores de MQO: Amostra Finita . . . . .	89
11.5.1	Estimadores não viesados . . . . .	89
11.5.2	Estimadores eficientes . . . . .	89
11.5.3	Inferência . . . . .	90
11.6	Propriedades assintóticas dos estimadores de MQO . . . . .	90
11.7	Modelos dinâmicos com variáveis não estacionárias . . . . .	91
11.7.1	Regressão espúria . . . . .	91

11.8 Cointegração . . . . .	93
11.9 Cointegração e modelos de correção de erro . . . . .	94
11.10 Estimando um modelo de correção de erro usando o método de dois passos de <i>Engle-Granger</i> . . . . .	95
11.11 Estimando um modelo de correção de erro em um passo . . . . .	100
11.12 Um pouco mais sobre cointegração . . . . .	102
11.12.1 Cointegração e Causalidade de <i>Granger</i> . . . . .	103
<b>Referências</b>	<b>105</b>

# 1 Introdução

A idéia deste trabalho é somente de fornecer uma "porta de entrada" para alunos com algum interesse em programação e em econometria de séries temporais. Vale destacar que ele se baseia fortemente no curso de séries temporais, dado no primeiro ano de mestrado em economia da Universidade de Estocolmo, com duração de 8 semanas.

Todo trabalho será feito no **GUI** do **R** chamado **RStudio**. Para começar você deve instalar o **R** e, em seguida, o **RStudio**. Caso tenha interesse também de aprender a produzir documentos  $\text{\LaTeX}$  utilizando o **RStudio** recomendo entrar em *Preferências* e em seguida *Sweave*. Finalmente, alterando a primeira opção disponível para **knitr**. Desta maneira a manipulação do código dentro do arquivo **.Rnw** será mais parecido com o próprio  $\text{\LaTeX}$ . Para isso será necessária a instalação do pacote *knitr* e do programa **MikTeX**.

Note que muitos dos pacotes não estarão disponíveis no seu **RStudio**, logo deverão ser instalados. Isto pode ser feito através do menu de *Ferramentas*.

Finalmente, note que você pode alterar o diretório de trabalho do **RStudio**, ou seja, onde todo o trabalho será salvo. Isto pode ser feito através do **RStudio** na janela inferior direita ou através do seguinte comando<sup>1</sup>:

```
setwd("C:/Desktop/Nome/R/")
```

## 1.1 Por que usar o R?

Nosso primeiro contato com econometria na prática é, quase sem exceção, através de programas econométricos de relativo fácil acesso a todos os comandos (como é o caso de *Gretl*, *Stata* e *Eviews*). Desta maneira, ficamos acostumados desde o início a clicar em determinados comandos e obter resultados diretamente, algo que didaticamente é ótimo especialmente em se tratando de uma matéria tão "adorada" como a econometria. Por isso, em geral, começar a usar o **R** (ou qualquer linguagem de programação) é um grande

---

<sup>1</sup>Note que se usa / e não \.

desafio pois além de aprender uma nova "língua" temos que aprender a nos expressar nela de maneira lógica. É esta "lógica" que na minha opinião é o mais difícil de se aprender.

Creio que o principal motivo de se utilizar o **R** está na capacidade de se repetir e reproduzir processos. É óbvio que podemos usar um destes programas para estimar 10-20 variações de um mesmo modelo, porém quanto tempo isso irá levar ? Como serão armazenados os resultados ?

Obviamente alguém pode então responder: mas também é possível programar e "automatizar" com estes outros programas! Sim, isto é verdade (apesar de muitos não saberem disso). O problema aqui é que em geral a linguagem é bastante específica para cada programa e, em geral, não seguem a lógica das linguagens mais comuns (como por exemplo Matlab).

De maneira mais direta aqui vai alguns pontos a favor e contra o uso do **R**:

## Usar **R**

- Gratuito;
- Não serve somente para econometria/estatística;
- Automatizar/repetir procedimentos;
- Similiaridade com outras linguagens de programação;
- Facilidade de divulgar/dividir resultados com outros programas e formatos<sup>2</sup> (L<sup>A</sup>T<sub>E</sub>X, Shiny, PDF, por exemplo);
- Pacotes e funções são atualizados constantemente;
- Mercado de trabalho (grandes empresas estão cada vez mais interessados em profissionais que saibam usar **R**);
- Te obriga a ter um conhecimento teórico maior;

## Não usar **R**

---

<sup>2</sup>Esta apostila, por exemplo, foi toda feita utilizando a interligação do **R** com o L<sup>A</sup>T<sub>E</sub>X.

- Custo alto de entrada;
- Trabalhoso para pequenos procedimentos;
- Econometria/estatística não faz parte do seu dia a dia;
- Te obriga a ter um conhecimento teórico maior;

## 1.2 O que é uma série de tempo?

Uma série de tempo é uma coleção de variáveis aleatórias ordenadas ao longo do tempo. O *PIB* brasileiro no intervalo entre 1950 e 2014, por exemplo, pode ser descrito como uma variável aleatória,  $Y_t$ , onde  $\{Y_t\}_{1950}^{2014} = \{Y_{1950}, Y_{1951}, \dots, Y_{2014}\}$ . Ou seja, cada  $Y_t$  é uma variável aleatória, porém somente observamos uma única realização dela.

Aqui nasce a principal diferença entre a econometria de *cross-section*, ou dados de corte, e a de séries temporais. Com dados de corte temos acesso a  $N$  observações das  $k$  variáveis de interesse para um dado instante no tempo, enquanto que no caso de séries de tempo somente temos 1 observação das  $k$  variáveis de interesse para um dado instante no tempo.

Veremos que esta pequena diferença traz grandes consequências para nosso objetivo final que, pelo menos na grande maioria dos casos, é fazer inferências sobre e tentar descrever o futuro desenvolvimento da variável de interesse.

## 1.3 Porque analisar séries de tempo?

Qual seria então nosso objetivo ao analisar dados de série de tempo? Basicamente o que queremos fazer é descrever estatisticamente os dados disponíveis e, além disso, descrevê-los matematicamente, ou seja, queremos uma equação que nos mostre como a série de interesse se comporta.

A partir disto podemos construir novas teorias econômicas ou simplesmente testar teorias já estabelecidas. Além disso, podemos, a partir do nosso conhecimento dos dados, tentar construir previsões para valores futuros das variáveis em mãos.

## 2 Começando os trabalhos

### 2.1 Importando e visualizando os dados

Existem muitas maneiras de importar dados para o **R**, sendo que os formatos mais fáceis de se trabalhar são *.txt* e *.csv*. A maneira mais simples de importar os dados para o **R** é usando o **RStudio** e clicando diretamente no botão "*Import Dataset*" no canto superior direito.

Como primeiro exemplo vamos começar com um banco de dados disponível no próprio **R** no pacote *pwt8*. O primeiro passo é instalar e carregar o pacote e, em seguida, devemos escolher os dados que vamos analisar:

```
library(pwt8) # Carregando o pacote
data("pwt8.0") # Carregando os dados do pacote

# Variáveis escolhidas são:
# Pib real
# Média de horas trabalhadas
# Taxa de c\^ambio

br <- subset(pwt8.0, country=="Brazil",
             select=c(rgdpna, avh, xr))
```

A partir desses dados podemos criar novas variáveis. Que tal criarmos uma *proxy*, definida como *PIB real* dividido pela *média de horas trabalhadas*, para produtividade:

```
br$prod <- br$rgdpna/br$avh
```

Para finalizar, podemos salvar os dados que separamos como um arquivo *.txt* e visualizar os dados através do comando *plot*.

```
write.table(br, file='dados.txt')
```

Agora temos o arquivo *dados.txt* no nosso diretório de trabalho. Podemos carregar este e, em seguida, visualizar os dados. Note que, na maioria das vezes, temos que avisar



ao **R** sobre o tipo de dados que estamos usando. Logo, neste caso devemos avisar que estamos utilizando séries de tempo.

```
dados <- read.table('dados.txt')
dados <- ts(dados, start=1950, freq=1)
```

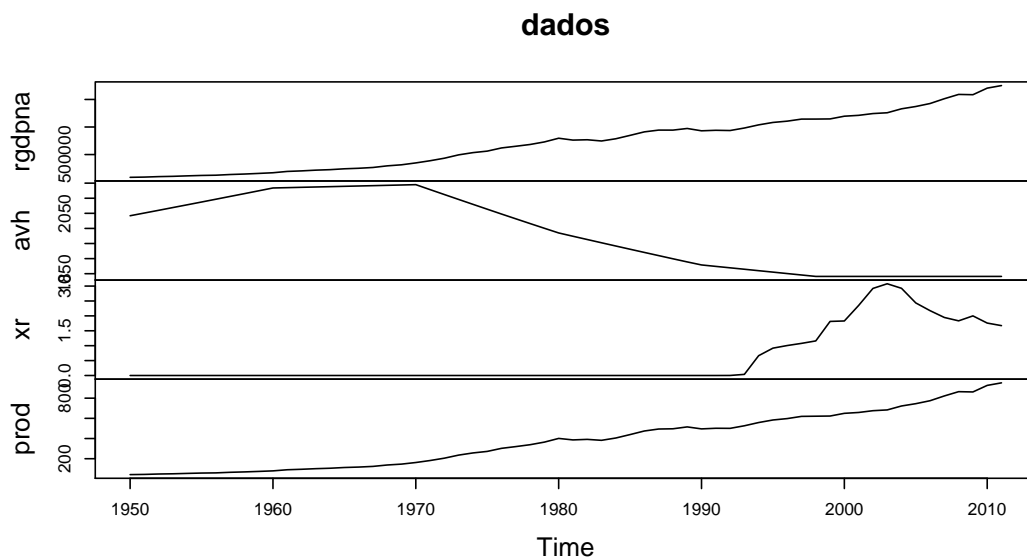
Aqui o comando *ts* está avisando ao **R** que a matriz *dados* possui séries temporais que se iniciam em 1950 com frequência 1, ou seja, são dados anuais. Frequência 4 e 12 seriam, respectivamente, dados trimestrais e mensais.

Note que poderíamos escrever este código de maneira mais curta :

```
dados <- ts(read.table('dados.txt'), start=1950, freq=1)
```

Para finalizar podemos visualizar os dados através do comando *plot*.

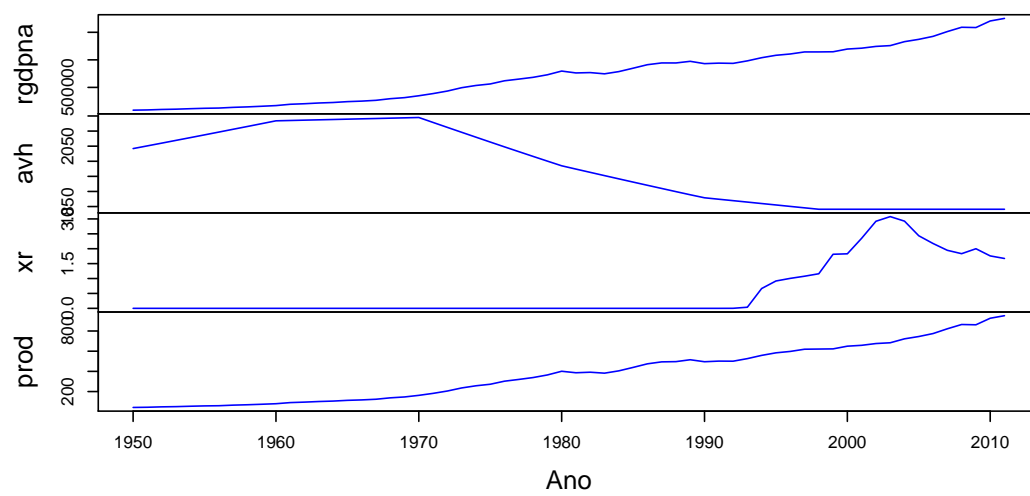
```
plot(dados)
```



Podemos aprimorar um pouco este gráfico através das opções do comando *plot*, como pode ser visto abaixo.

```
plot(dados,
     col='blue',
     main='Dados brasileiros',
     xlab='Ano')
```

## Dados brasileiros



### 3 Fatos estilizados do Ciclo de Negócios

Quais são então os fatos que devemos buscar entender quando queremos estudar as variáveis macroeconômicas e, principalmente, os ciclos econômicos? As principais características que precisamos são a *volatilidade*, *correlação contemporânea* e se a variável em questão é *antecedente* ou não ao desenvolvimento do PIB.

Que tal vermos o caso brasileiro utilizando os dados do *Banco Central do Brasil*?

	Volatilidade Rel.	Volatilidade Abs.	(t-1)	(t)	(t+1)
PIB	1.00	1.43	0.66	1.00	0.66
CF	1.10	1.57	0.43	0.67	0.57
CG	0.90	1.88	-0.10	0.19	0.14
I	1.37	5.08	0.74	0.91	0.58

Tabela 1: Comportamento cíclico da economia brasileira.

Como podemos ver da tabela anterior...

#### 3.1 Removendo tendências de séries univariadas

O motivo pelo qual queremos remover tendências de nossas variáveis macroeconômicas, pelo menos neste contexto, é de separar tendências de crescimento de longo prazo e variações sazonais de fenômenos exclusivamente cíclicos, ou aleatórios. Ao longo desta seção tenha em mente a seguinte equação:

$$X_t = sazonalidade_t + tendencia_t + ciclo_t + \varepsilon_t$$

##### 3.1.1 Removendo sazonalidade

A maioria das séries macroeconômicas apresentam algum tipo de sazonalidade e, como nós estamos interessados no comportamento cíclico de nossas variáveis devemos tentar "limpar" nossas séries desse componente sazonal antes de tudo.

Existem muitas maneiras de se tentar eliminar a sazonalidade dos dados, mas para o nosso caso usaremos o método mais simples, através de *dummies*. Repare que esse método

somente removerá a parte *determinística* da sazonalidade<sup>3</sup>.

Assuma agora que temos dados trimestrais, assim podemos fazer a seguinte regressão:

$$Y_t = \alpha + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \varepsilon_t$$

Repare que incluímos *dummies* para cada trimestre exceto o primeiro, isto para não termos problemas de linearidade perfeita. Isto também significa que todos os efeitos medidos por  $D_i$  serão relativos ao primeiro trimestre.

Feita a regressão podemos efetuar um *teste F*, onde  $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ , ou seja, testamos se todos os coeficientes são estatisticamente conjuntamente iguais a zero. Isto é, estamos testando a presença de uma sazonalidade trimestral determinística nos dados.

Se rejeitarmos a hipótese nula estaremos aceitando a hipótese de que a série  $Y_t$  apresenta sazonalidade. Logo, podemos usar  $\hat{\varepsilon}_t$  como nossa série dessazonalizada. Veja porque,

$$Y_t = \hat{\alpha} + \hat{\beta}_2 D_2 + \hat{\beta}_3 D_3 + \hat{\beta}_4 D_4 + \hat{\varepsilon}_t$$

$$Y_t = \hat{Y}_t + \hat{\varepsilon}_t$$

Logo, temos que a própria série menos a parte que capta os efeitos da sazonalidade é igual a parte "filtrada" da série, ou seja, a série dessazonalizada.

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t$$

Obviamente, se não rejeitarmos a hipótese nula teríamos que a própria série original representaria a série dessazonalizada.

Agora vamos para um exemplo prático. Neste próximo exercício usaremos o PIB trimestral (série 1232 do *SGS* do Banco Central do Brasil). Como sempre começamos importando os dados e, em seguida, visualizando eles.

```
pib <- ts(read.delim("pib.csv", dec=","),
          start=1991, freq=4)

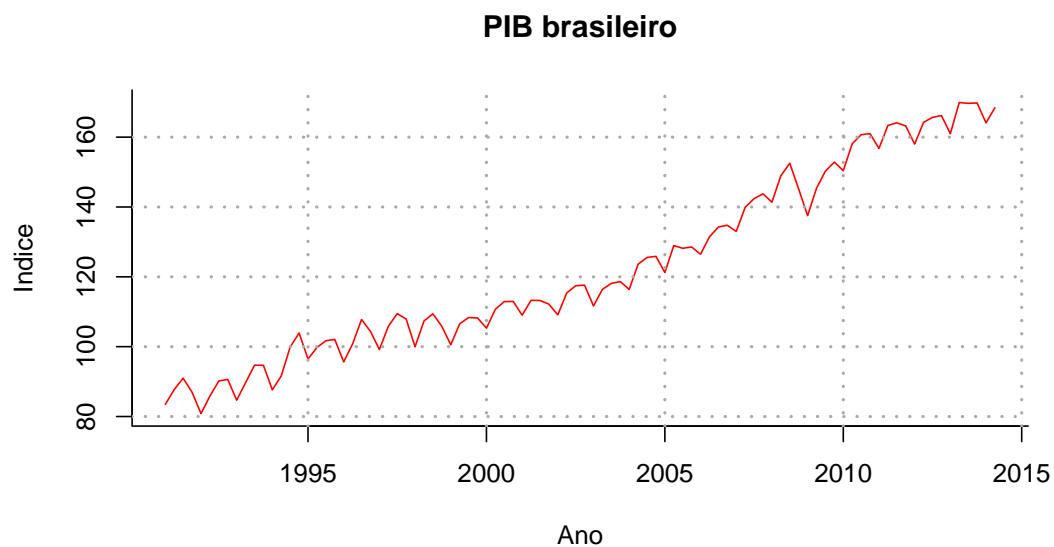
pib <- pib[,-1]
```

---

<sup>3</sup>Sazonalidade estocástica será abordado mais à frente.

```
plot(pib,main='PIB brasileiro',
     ylab='Indice', xlab = 'Ano',
     bty='l',col='red',lty=1)

grid(col='darkgrey',
     lwd=2)
```



Agora devemos rodar a regressão, para isso devemos criar os *dummies* sazonais.

```
Q <- ordered(cycle(pib))

pib.reg <- lm(pib~Q)
```

Utilizando o comando *summary* podemos obter um resumo da regressão.

```
summary(pib.reg)

##
## Call:
## lm(formula = pib ~ Q)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -38.031 -18.642 -7.356 21.339 46.177
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 122.6690     2.7099  45.266  <2e-16 ***
## Q.L          4.4144     5.4199   0.814   0.418
## Q.Q         -3.2150     5.4199  -0.593   0.555
## Q.C          0.8823     5.4199   0.163   0.871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.27 on 90 degrees of freedom
## Multiple R-squared:  0.01159, Adjusted R-squared:  -0.02136
## F-statistic: 0.3516 on 3 and 90 DF,  p-value: 0.7881
```

Como podemos ver, o resumo já nos dá o p-valor do teste  $F$  que neste caso é igual a 0.79<sup>4</sup>. Logo, não rejeitamos a hipótese nula de que não existe sazonalidade. Porém, para fins didáticos vamos assumir que rejeitamos a  $H_0$  e continuaremos com o procedimento de dessazonalizar os dados.

Agora, olhemos para os resíduos da regressão, ou seja, nossa série dessazonalizada, e o componente sazonal.

```
pib.des <- ts(resid(pib.reg),
              start = 1991, freq=1)

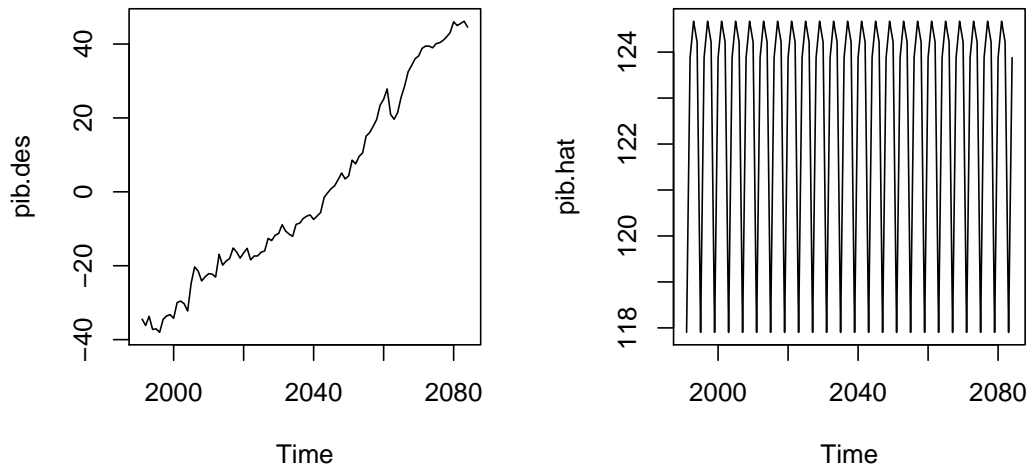
pib.hat <- ts(fitted(pib.reg),
              start=1991, freq=1)

par(mfrow=c(1,2))
plot(pib.des)
```

---

<sup>4</sup>Note que nem individualmente, através dos testes  $t$  nenhum dos componentes sazonais são estatisticamente significativos.

```
plot(pib.hat)
```



Como podemos ver, a série dessazonalizada começa em níveis negativos. Isto acontece pois ao subtrair  $\hat{Y}_t$  da nossa série original estamos também retirando a sua média, captada por  $\hat{\alpha}$ . Logo, devemos "normalizar" nossa série adicionando sua média.

```
pib.desn <- pib.des + mean(fitted(pib.reg))
```

Agora podemos fazer um gráfico com a série original e sua versão dessazonalizada.

```
par(mfrow=c(1,1))

plot(pib,
     main='',
     xlab='Ano', ylab='',
     col='blue',
     bty='l')

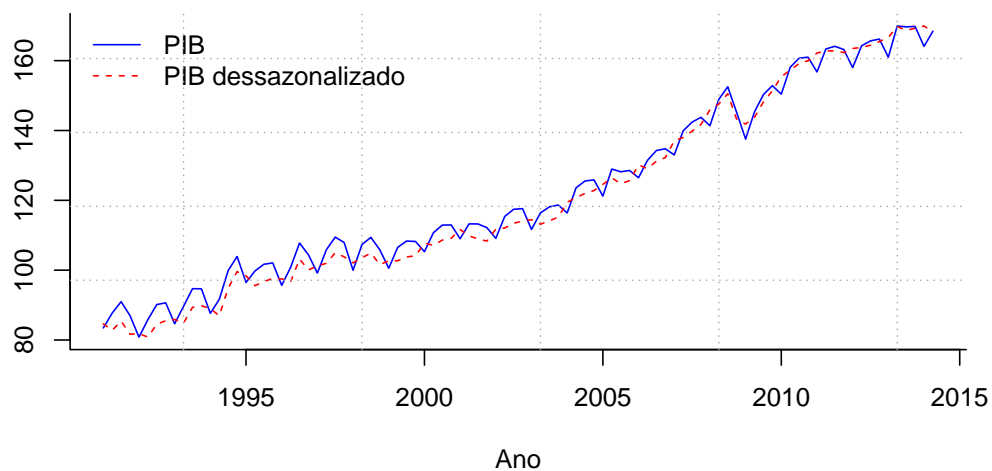
par(new=TRUE)

plot(pib.desn,
     axes=F, ann=F,
     col='red',
```

```
lty=2)

legend('topleft',
      c('PIB', 'PIB dessazonalizado'),
      col=c('blue', 'red'), lty=1:2,
      bty='n')

grid(col='darkgrey')
```



### 3.1.2 Tendência de longo prazo

Assumindo agora que nós temos dados anuais podemos rodar a seguinte regressão:

$$PIB_t = \alpha + \beta Ano_t + \varepsilon_t$$

Aqui, assim como no caso da sazonalidade, se  $\hat{\beta}$  for estatisticamente diferente de zero podemos concluir que a série em questão não possui uma tendência linear. Desta forma, se o valor de  $t$  for estatisticamente significativo a série  $\hat{\varepsilon}_t$  representará  $PIB_t$  removido de sua tendência linear. Caso contrário, podemos utilizar a série original.

Para este próximo exercício iremos utilizar a série 1207 do Banco Central do Brasil que nos dá o PIB em reais correntes desde 1962.

De maneira semelhante ao que fizemos antes iremos importar os dados, avisar ao **R** que se trata de séries de tempo e, em seguida, rodaremos a regressão descrita acima. Note



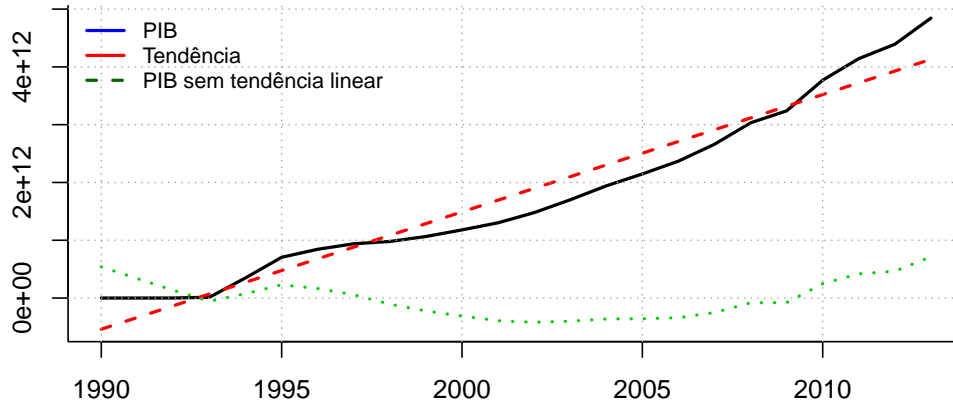
que para isso precisamos criar um vetor representando a tendência linear (por exemplo uma sequência de 1962 até 2013, ou simplesmente de 1 até 52).

Os resultados da regressão podem ser vistos a seguir:

```
##
## Call:
## lm(formula = pibanual ~ ano)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.194e+11 -3.188e+11 -6.747e+10  2.344e+11  7.142e+11
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.046e+14  2.020e+13  -20.03 1.29e-15 ***
## ano          2.030e+11  1.009e+10   20.12 1.17e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.422e+11 on 22 degrees of freedom
## Multiple R-squared:  0.9485, Adjusted R-squared:  0.9461
## F-statistic: 404.8 on 1 and 22 DF,  p-value: 1.175e-15
```

Como podemos ver o  $p$ -valor do teste do qual temos interesse é igual a 0, logo podemos rejeitar a hipótese de que a série não apresenta uma tendência linear. Isto significa que  $\hat{\varepsilon}_t$  representa nossa série sem tendência, enquanto que  $\widehat{PIB}_t$  será a nossa tendência linear.

Em seguida temos as três séries no mesmo gráfico.



Obviamente, uma série não necessariamente possui somente uma tendência linear. Assuma, por exemplo, que a série de nosso interesse siga o seguinte processo

$$y_t = \alpha + \beta_1 ano_t + \beta_2 ano_t^2 + \varepsilon_t$$

Neste caso será então necessário remover a parte quadrática da tendência.

### Filtro Hodrick-Prescott

O filtro *Hodrick-Prescott*, ou simplesmente filtro HP, é um método comumente utilizado na hora de retirar tendências de longo prazo de séries macroeconômicas. Este permite a existência de uma tendência mais "flexível" dado que ele não assume uma tendência perfeitamente linear.

Considere  $X_t$  como a soma de um componente cíclico,  $X_t^c$ , um componente de crescimento,  $X_t^g$ , e um choque aleatório contemporâneo,  $\varepsilon_t$ . Logo, podemos representar  $X_t$  como:

$$X_t = X_t^c + X_t^g + \varepsilon_t$$

Seja  $\lambda$  um parâmetro que representa a variância relativa do componente de crescimento ao componente cíclico. Agora, dado um valor de  $\lambda$ , deve-se escolher um componente de crescimento que minimize a seguinte função perda:

$$\Lambda = \frac{1}{T} \sum_{t=1}^T (X_t - X_t^g)^2 + \frac{\lambda}{T} \sum_{t=1}^T [(X_{t+1}^g - X_t^g) - (X_t^g - X_{t-1}^g)]^2$$

Note que quanto maior for o valor de  $\lambda$  maior será a penalização no termo de taxa de variação do componente de crescimento. Logo, quanto maior o valor de  $\lambda$  mais  $X_t^g$  se aproximará de uma tendência linear. Caso contrário, ou seja, quando  $\lambda$  se aproxima de zero, teremos que  $X_t^g = X_t$ .

Abaixo criamos uma função que calcula o filtro HP para uma série.

```
filtrohp <- function(y, lambda){  
  
  id <- diag(length(y))  
  d <- diff(id, d=2)  
  tendhp <- solve(id+lambda*crossprod(d), y)  
  tendhp  
  
}
```

Em geral, os valores recomendados<sup>5</sup> de  $\lambda$  são aproximadamente 100 para dados anuais, 1600 para dados trimestrais e 14400 para dados mensais.

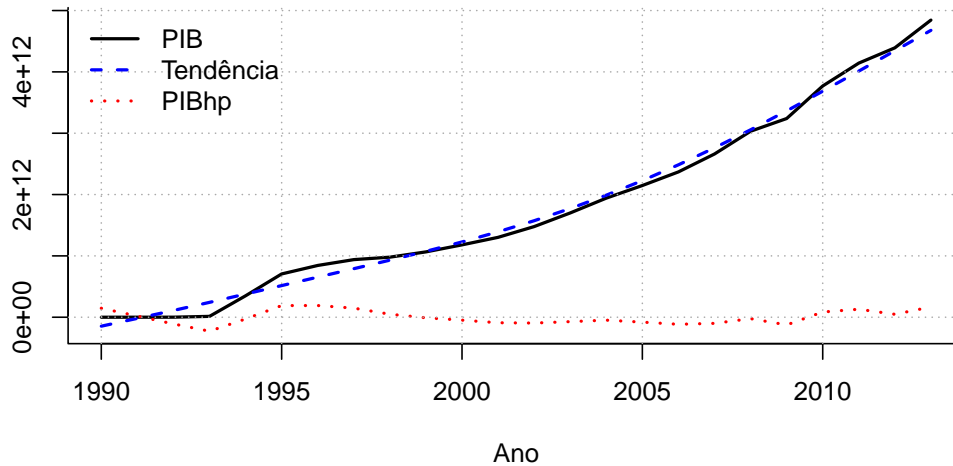
Agora podemos prosseguir e criar a nossa variável sem tendência através do uso do filtro HP.

```
lambda <- 100  
tendencia <- filtrohp(pibanual, lambda)  
hppib <- pibanual - tendencia
```

Finalmente, visualizamos os dados.

---

<sup>5</sup>Note que os valores recomendados dependem fortemente dos autores.



## Diferenças

Tomar a primeira diferença é outra maneira simples de se remover a tendência de uma série de tempo. Mais a frente usaremos muito esta ferramenta, porém neste momento talvez não seja o jeito mais consistente teoricamente para remover a tendência de uma série. Apesar de alguns ainda utilizarem este método, um método mais popular seria de tomar o logaritmo natural e, em seguida, diferenci

No **R** podemos tomar a primeira de uma série através do comando *diff*. Com esta função podemos usar diferenças até maiores, ou seja, podemos tomar a  $n$ -ésima diferença. Abaixo estão alguns exemplos de como usar este comando:

- A primeira diferença da série  $\{x_t\}$  é calculada como  $x_t - x_{t-1} \quad \forall \quad t$ . Para obter isso no **R** digite `diff(x, lag=1, differences=1)` ou, de maneira mais simples, `diff(x)`;
- A segunda diferença de  $\{x_t\}$  é calculada como  $[(x_t - x_{t-1}) - (x_{t-1} - x_{t-2})] \quad \forall t$ . No **R** podemos escrever `diff(x, 1, 2)`;
- A primeira diferença sazonal, no caso em que  $\{x_t\}$  representa uma série trimestral, é calculada como  $x_t - x_{t-4} \quad \forall t$ . No **R** obtemos isso ao digitar `diff(x, 1, 4)` ou simplesmente `diff(x, 4)`.

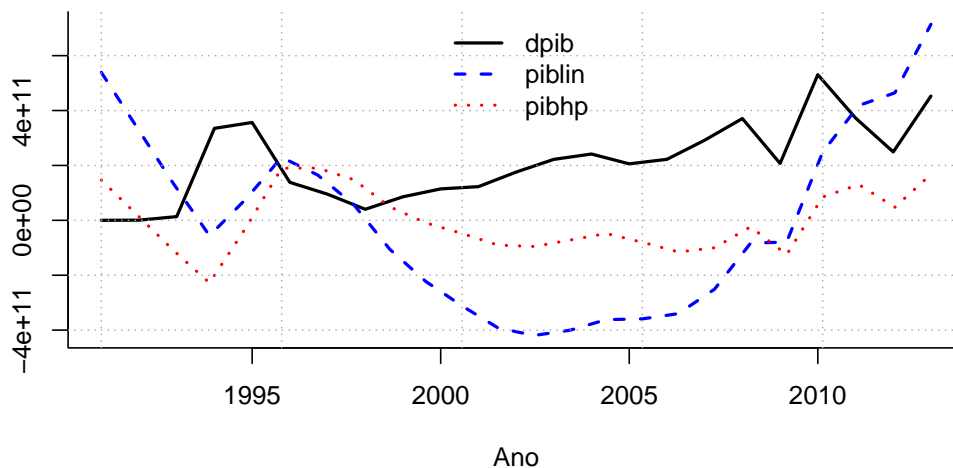
## Comparação entre os três métodos de remoção de tendência

Agora que estamos um pouco mais familiarizados com estas três maneiras de retirar tendências de nossos dados, que tal compararmos visualmente a diferença? As séries são `piblin`, `pibhp` e `dpib`, onde a primeira representa a remoção da tendência linear, a segunda a remoção através do filtro HP e a última representa a remoção através da primeira diferença.

```
dpib <- diff(pibanual,1)

piblin <- ts(resid(lm(pibanual~seq(1,length(pibanual)))),
             start=1990, freq=1)

pibhp <- ts(pibanual - filtrohp(pibanual, 100),
            start=1990, freq=1)
```



Podemos examinar a correlação contemporânea dessas nossas variáveis através do comando `cor(x,y)`.

```
cor(piblin, pibhp)

## [1] 0.6934064
```

```
cor(piblin[-1], dpib)
```

```
## [1] 0.3378531
```

```
cor(pibhp[-1], dpib)
```

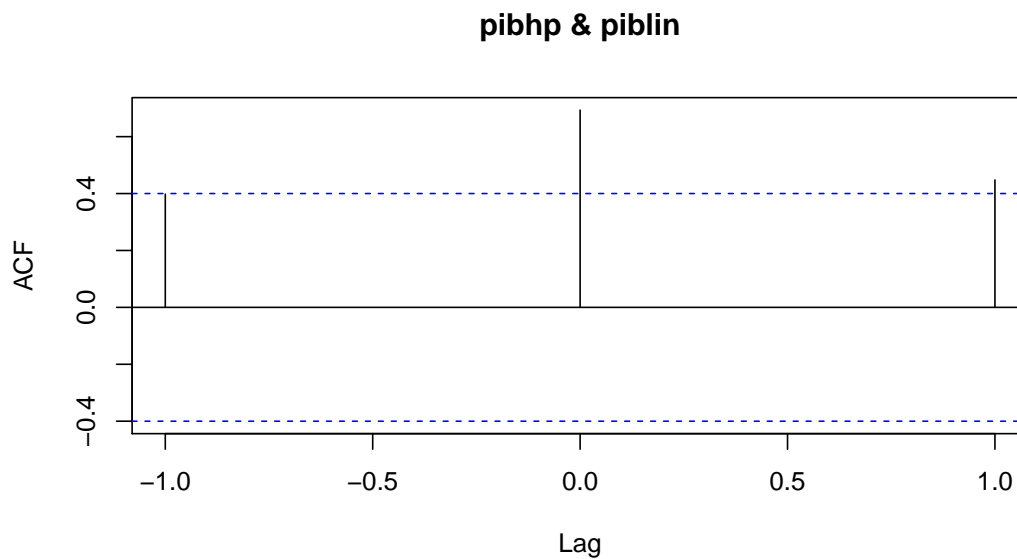
```
## [1] 0.3861809
```

Como podemos ver as três séries são bastantes distintas entre si, porém isto não significa que os fatos estilizados das três séries sejam necessariamente estatisticamente diferentes. Você deveria extrair as fatos estilizados utilizando a série de cada um desses métodos e observar a diferença entre eles. Assim, estará estudando o impacto que cada um deste métodos têm no conjunto de fatos.

### 3.2 Calculando correlações cíclicas

Como foi mostrado anteriormente, para calcular a correlação contemporânea entre duas séries  $x_t$  e  $y_t$  podemos simplesmente escrever `cor(x,y)`. Porém, muitas vezes estamos interessados na correlação entre duas variáveis com uma defasagem maior entre elas. Neste caso podemos utilizar a função de correlação cruzada do **R** chamada `ccf`, que além de calcular as correlações cruzadas e seus respectivos intervalos de confiança também nos fornece um gráfico. Vejamos um exemplo para calcular a correlação de `piblin(t)` com `pibhp(t)`, `pibhp(t-1)` e `pibhp(t+1)`.

```
cc <- ccf(pibhp, piblin, 1)
```



```
cc

##
## Autocorrelations of series 'X', by lag
##
##      -1      0      1
## 0.398 0.693 0.448
```

Note que a linha azul nos dá o intervalo de confiança de 95%, ou seja, dentro deste intervalo estão os valores que são estatisticamente não significativos.

### 3.3 Defasagens no R

Para defasar uma série  $x_t$   $n$  vezes utilizamos `lag(x, k=-n)`. Da mesma forma, para "avançar" uma série  $n$  vezes utilizamos o mesmo comando porém com o positivo de  $n$ , ou seja, `lag(x, k=n)`. De modo a ilustrar podemos fazer um gráfico com as diferentes séries onde utilizamos  $n = 5$ .

```
ts.plot(pibanual, lag(pibanual, 5), lag(pibanual,-5),
        col=1:3,
        xlab='Ano', main='Demonstrando defasagens',
        lty=1:3, lwd=2)
```

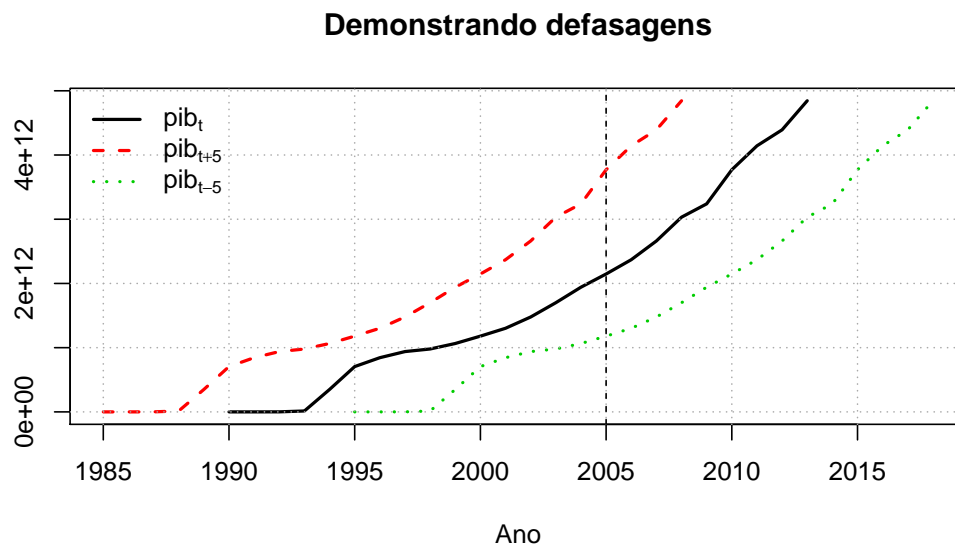
```

grid(col='darkgrey')

legend('topleft',
      c(expression(pib[t]), expression(pib[t+5]), expression(pib[t-5])),
      col=1:3, lty=1:3,
      bty='n', lwd=2)

abline(v=2005, lty=2)

```



Repare que ao traçarmos uma linha vertical no ano de 2005, por exemplo, estaremos atingindo os valores de 2000, 2005 e 2010, respectivamente. Isto significa que se estivermos interessados em rodar uma regressão do PIB na sua segunda defasagem poderíamos utilizar os seguintes comandos:

```
reg.pib <- lm(pibanual~lag(pibanual, -2))
```

### 3.4 Calculando a volatilidade cíclica

A proposta deste seguinte exercício é a de determinar a volatilidade da variável  $x$ , ou seja da magnitude das flutuações em  $x$ , decorrente de seu próprio comportamento cíclico. Aqui



somente duas<sup>6</sup> maneiras de representar esta volatilidade serão abordadas, a volatilidade *absoluta* e a volatilidade *relativa*. A primeira mede a "própria" volatilidade da variável como o desvio padrão (d.p.) da flutuação percentual em torno da tendência de longo prazo. Esta volatilidade pode ser calculada como a seguir

$$\text{volatilidade } x = d.p. \left[ \frac{x_t - \text{tendencia}(x)_t}{\text{tendencia}(x)_t} * 100 \right].$$

Por outro lado, a volatilidade relativa de  $x$  é sua volatilidade relativa a uma outra série. Por exemplo, a volatilidade de  $x$  relativa a uma outra variável,  $y$ , pode ser calculada como

$$\text{volatilidade } x/y = \frac{d.p.(\ln x)}{d.p.(\ln y)}.$$

Agora que sabemos calcular os diferentes tipos de volatilidade podemos reproduzir os dados da Tabela 1. Para o caso da volatilidade absoluta do PIB, por exemplo, deveríamos seguir o código a seguir<sup>7</sup>.

```
vol.pib <- sd(((pib - filtrohp(pib, 1600))/(filtrohp(pib, 1600)))*100)
vol.pib

## [1] 2.969671
```

---

<sup>6</sup>Existem muitas maneiras diferentes de se medir volatilidade, por isso sempre é importante ter em mente qual definição você está usando e sempre deixar de maneira explícita.

<sup>7</sup>Observe que neste exemplo estamos calculando a volatilidade do série sem nenhum tratamento, ou seja, ainda não foi dessazonalizada. Logo, o valor pode ser diferente ao da tabela.

## 4 Modelos univariados

Suponha que temos o seguinte modelo geral univariado de séries temporais

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots, u_t)$$

De modo a operacionalizar este modelo temos que especificar três coisas;

1. A forma funcional de  $f(\cdot)$ ;
2. O número de defasagens;
3. A estrutura do termo de distúrbio,  $u_t$ .

A base lógica para a análise univariada pode ser resumida, de certa forma, nos seguintes pontos

1. Modelos puramente estatísticos (ateóricos) podem ser extremamente úteis quando queremos resumir informações/características de uma série de tempo e para fazer previsões bastante confiáveis de curto prazo.
2. Sempre visualize os dados ANTES de rodar as regressões! Muitas vezes os gráficos já nos dizem bastante sobre a natureza do processo por trás dos dados.
  - (a) Existe algum padrão de sazonalidade?
  - (b) Há alguma tendência de longo prazo?
  - (c) Existem quebras estruturais e/ou eventos históricos atípicos ?
3. Modelos teóricos com defasagens da variável dependente podem muitas vezes serem reduzidos a modelos univariados relevantes.
4. Modelos de séries temporais são importantes para métodos teóricos, analíticos e numéricos. Não somente para econometria.

**Exemplo:** Um modelo macroeconômico simples

$$c_t = \beta_0 + \beta_1 y_t + \beta_2 c_{t-1} + \varepsilon_t \tag{1}$$

$$y_t = c_t + i_t \tag{2}$$

$$i_t = sy_t \quad (3)$$

onde  $0 < s < 1$  é uma taxa de poupança constante,  $i_t$  é o investimento,  $c_t$  é o consumo,  $y_t$  é a renda e  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$  representa um ruído branco. Substituindo (3) em (2) temos que

$$c_t = \beta_0 + \beta_1 y_t + \beta_2 c_{t-1} + \varepsilon_t \quad (4)$$

$$y_t = c_t + sy_t. \quad (5)$$

Logo, substituindo (5) em (4)

$$c_t = \beta_0 + \beta_1 \frac{c_t}{1-s} + \beta_2 c_{t-1} + \varepsilon_t$$

$$c_t - \frac{\beta_1}{1-s} c_t = \left(1 - \frac{\beta_1}{1-s}\right) c_t = \beta_0 + \beta_2 c_{t-1} + \varepsilon_t$$

$$c_t = \alpha_0 + \alpha_1 c_{t-1} + u_t \quad (6)$$

onde  $u_t \sim N(0, \sigma_u^2)$ . Desta forma reduzimos o modelo macroeconômico para um processo univariado, ou seja, um processo  $A(1)$  para o consumo.

## 4.1 Alguns conceitos básicos

Começamos agora a definir alguns conceitos básicos que iremos usar extensivamente ao longo do curso.

**Processo estocástico** é uma coleção de variáveis aleatórias ordenadas no tempo. Por exemplo, tome o PIB brasileiro de 1950 até 2013  $\equiv \{Y_t\}_{1950}^{2013} = \{Y_{1950}, Y_{1951}, \dots, Y_{2013}\}$ . Cada  $Y_t$  é uma variável aleatória (isto é, cada  $Y_t$  é uma realização de um número infinito de "realidades" alternativas). O processo estocástico  $\{Y_t\}_{1950}^{2013}$  nos ajuda a descrever e inferir sobre o desenvolvimento do PIB brasileiro ao longo do tempo.

**Estacionaridade estrita** implica na distribuição de probabilidade conjunta constante ao longo do tempo, ou seja, ela não depende do instante do  $t$ .

**Estacionaridade fraca** implica que os dois primeiros momentos (ou seja, a média e a matriz de variância-covariância) da distribuição de probabilidade conjunta da variável estocástica sejam constantes ao longo do tempo, isto é

$$E\{Y_t\} = \mu < \infty$$

$$V\{Y_t\} = E\{(Y_t - \mu)^2\} = \gamma_0 < \infty$$

$$Cov\{Y_t, Y_{t-k}\} = E\{(Y_t - \mu)(Y_{t-k} - \mu)\} = \gamma_k < \infty, \forall k \in \mathbb{N}^+$$

Por agora, o conceito de estacionaridade fraca (de agora em diante *estacionaridade*) será suficiente para o nosso trabalho. Esta condição supre a maioria de nossas necessidades (i.e., ela nos permite usar as ferramentas padrões de estimação, previsão e inferência).

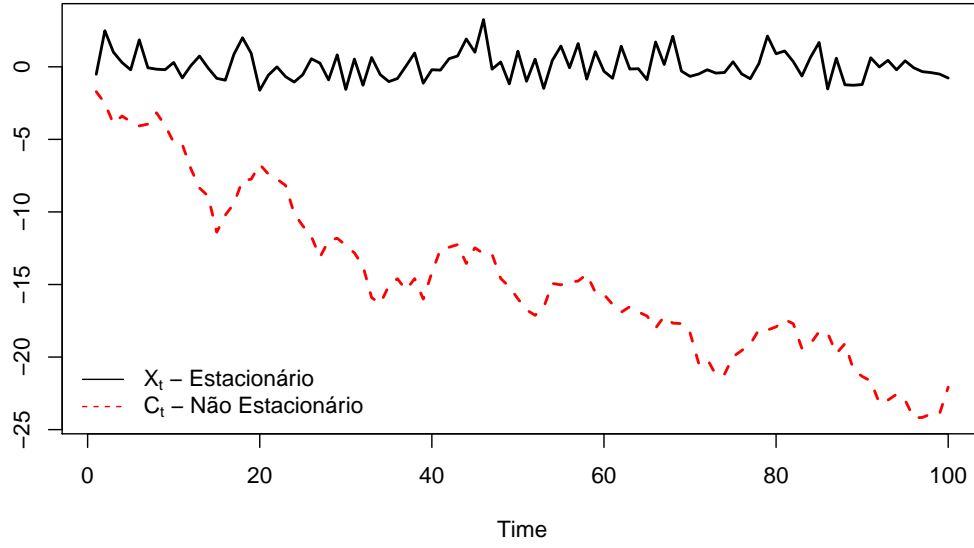
**Processo não estacionário** é um processo cuja média e variância serão dependentes do tempo. O fato desses dois momentos serem funções do tempo implica na impossibilidade de fazermos inferência fora da amostra.

**Processos puramente aleatórios**, ou também "ruídos brancos", são processos que possuem média zero, variância constante e não possuem correlação serial. Frequentemente queremos assumir que os choques nos nossos modelos econométricos são processos puramente aleatórios (obviamente iremos testar se esta hipótese válida ou não), isto é, iremos assumir que

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

Como exemplo podemos agora criar dois processos estocásticos, um estacionário e outro não estacionário.

```
set.seed(22)
TT <- 100
Xt <- ts(rnorm(TT), start=1, freq=1)
Ct <- ts(cumsum(rnorm(TT)), start=1, freq=1)
```



## 4.2 Processos de médias móveis

Um processo  $MA(1)$  pode ser caracterizado como

$$Y_t = \mu + \varepsilon_t + \alpha\varepsilon_{t-1}$$

onde

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2).$$

A média da série  $Y_t$  pode ser calculada e é igual a

$$E\{Y_t\} = E\{\mu + \varepsilon_t + \alpha\varepsilon_{t-1}\} = \mu.$$

Calculando a variância temos

$$\begin{aligned} V\{Y_t\} &= E\{(Y_t - \mu)^2\} = E\{(\varepsilon_t + \alpha\varepsilon_{t-1})^2\} \\ &= E\{\varepsilon_t^2 + \alpha^2\varepsilon_{t-1}^2 + 2\varepsilon_t\alpha\varepsilon_{t-1}\} = E\{\varepsilon_t^2\} + \alpha^2E\{\varepsilon_{t-1}^2\} + 2\alpha E\{\varepsilon_t\varepsilon_{t-1}\} \\ &= E\{\varepsilon_t^2\} + \alpha^2E\{\varepsilon_{t-1}^2\} \\ &= (1 - \alpha^2)\sigma_\varepsilon^2. \end{aligned}$$

A autocovariância entre  $Y_t$  e  $Y_{t-1}$  é dada por

$$\text{cov}\{Y_t, Y_{t-1}\} = E\{(Y_t - \mu)(Y_{t-1} - \mu)\}$$

$$\begin{aligned}
&= E\{(\varepsilon_t + \alpha\varepsilon_{t-1})(\varepsilon_{t-1} + \alpha\varepsilon_{t-2})\} \\
&= E\{(\alpha\varepsilon_{t-1}^2)\} \\
&= \alpha\sigma^2
\end{aligned}$$

Finalmente, a autocovariância entre  $Y_t$  e  $Y_{t-2}$  é dada por

$$\begin{aligned}
cov\{Y_t, Y_{t-2}\} &= E\{(Y_t - \mu)(Y_{t-2} - \mu)\} \\
&= E\{(\varepsilon_t + \alpha\varepsilon_{t-2})(\varepsilon_{t-2} + \alpha\varepsilon_{t-3})\} \\
&= 0
\end{aligned}$$

Note que para este processo  $MA(1)$ , teremos que  $cov\{Y_t, Y_{t-k}\} = 0 \quad \forall k > 1$ . Dito isto, como será a matriz variância-covariância,  $\sum_{\sigma}$ , deste processo?

No **R** podemos simular e visualizar um processo  $MA(1)$  como a seguir

```
plot(arima.sim(model=list(ma=0.8), n=100),
     ylab=expression(Y[t]),
     main='',
     bty='l')
```

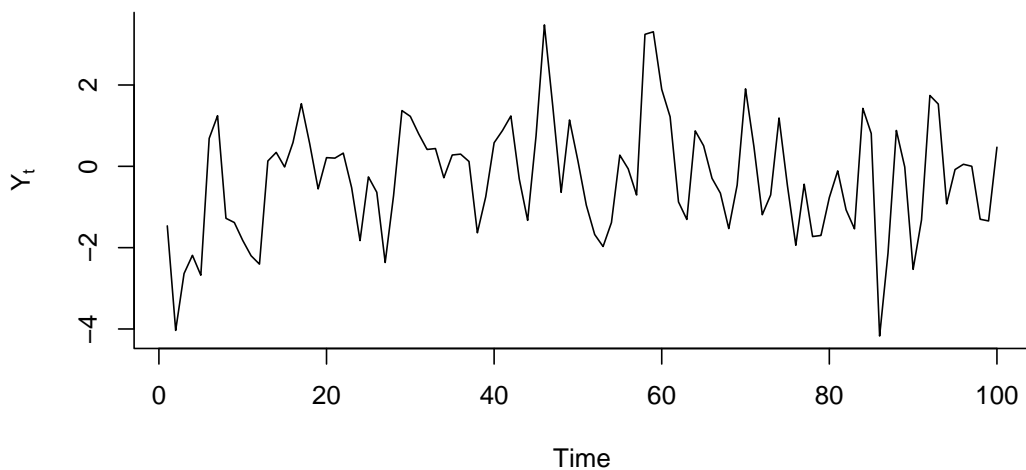


Figura 1: Processo  $MA(1)$  com  $\alpha = 0.8$ .

Veremos mais a diante que se  $|\alpha| < 1$ , este processo  $MA(1)$  pode ser invertido e transformado em um processo autoregressivo infinito, ou  $AR(\infty)$ . Em termos mais gerais podemos definir um processo  $MA(q)$  como

$$Y_t = \mu + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}$$

onde, como de costume,

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2).$$

A condição que garante a invertibilidade de um processo  $MA(q)$  é que o inverso das raízes do polinômio estejam dentro do círculo unitário.

### 4.3 Processos autoregressivos

O exemplo mais simples de um processo autoregressivo é dado abaixo

$$Y_t = \mu + \theta Y_{t-1} + \varepsilon_t$$

onde, mais uma vez, assumimos que

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2).$$

Calculando a esperança deste processo temos

$$E\{Y_t\} = E\{\mu + \theta Y_{t-1} + \varepsilon_t\} = \mu + E\{\theta Y_{t-1}\}$$

onde, se a nossa hipótese de estacionaridade é válida, temos que

$$E\{Y_t\} = E\{Y_{t-1}\} = E\{Y_{t-k}\} \quad \forall k$$

logo,

$$E\{Y_t\} = \mu + E\{\theta Y_t\}$$

$$(1 - \theta)E\{Y_t\} = \mu$$

$$E\{Y_t\} = \frac{\mu}{1 - \theta}.$$

Se lembrarmos das propriedades de uma P.G. infinita podemos ver que esta última expressão pode ser reescrita como

$$E\{Y_t\} = \mu(1 + \theta + \theta^2 + \dots + \theta^\infty) \iff |\theta| < 1.$$

**Teorema Séries Geométricas:** Se  $|r| < 1$ , a série geométrica  $a + ar + ar^2 + \dots + ar^n + \dots$  converge para  $a/(1 - r)$  quando  $n \rightarrow \infty$ . Se  $|r| \geq 1$ , a série divergirá caso  $a \neq 0$ . Caso contrário, a série convergirá para 0.

De forma a facilitar os cálculos da variância deste processo  $AR(1)$ , redefina  $y_t = Y_t - \mu$ , logo

$$\begin{aligned} V\{y_t\} &= E\{(y_t - E\{y_t\})^2\} = E\{(\theta y_{t-1} - \varepsilon_t)^2\} \\ &= E\{(\theta^2 y_{t-1}^2 - 2\theta y_{t-1} \varepsilon_t + \varepsilon_t^2)\} = \theta^2 E\{y_{t-1}^2\} + E\{\varepsilon_t^2\} \\ &= \sigma_\varepsilon^2 + \theta^2 E\{(\theta y_{t-2} + \varepsilon_{t-2})^2\} = \dots \\ &= \sigma_\varepsilon^2(1 + \theta^2 + \theta^4 + \dots + \theta^\infty) \\ &= \frac{\sigma_\varepsilon^2}{1 - \theta^2}. \end{aligned}$$

Ainda tomando  $y_t = Y_t - \mu$ , a autocovariância pode ser calculada como a seguir

$$\begin{aligned} Cov\{y_t, y_{t-1}\} &= E\{(y_t - E\{y_t\})(y_{t-1} - E\{y_{t-1}\})\} = E\{y_t y_{t-1}\} \\ &= E\{(\theta y_t + \varepsilon_t) y_{t-1}\} = E\{\theta y_{t-1} y_{t-1} + \varepsilon_{t-1}\} \\ &= \theta E\{y_{t-1}^2\} = \theta V\{y_t\} \\ &= \frac{\theta \sigma_\varepsilon^2}{1 - \theta^2} \iff |\theta| < 1. \end{aligned}$$

Generalizando, teremos que a autocovariância entre  $y_t$  e  $y_{t-k}$  será dada pela expressão

$$cov\{y_t, y_{t-k}\} = \frac{\theta^k \sigma_\varepsilon^2}{1 - \theta^2} \iff |\theta| < 1.$$

De modo a visualizar um processo  $AR(1)$  podemos seguir os códigos a seguir.

```
Yt <- arima.sim(model=list(ar=0.8), n=100)
plot(Yt,
      lty=1, bty='l', col=1,
      main='',
      ylab=expression(paste(Y[t])))
```



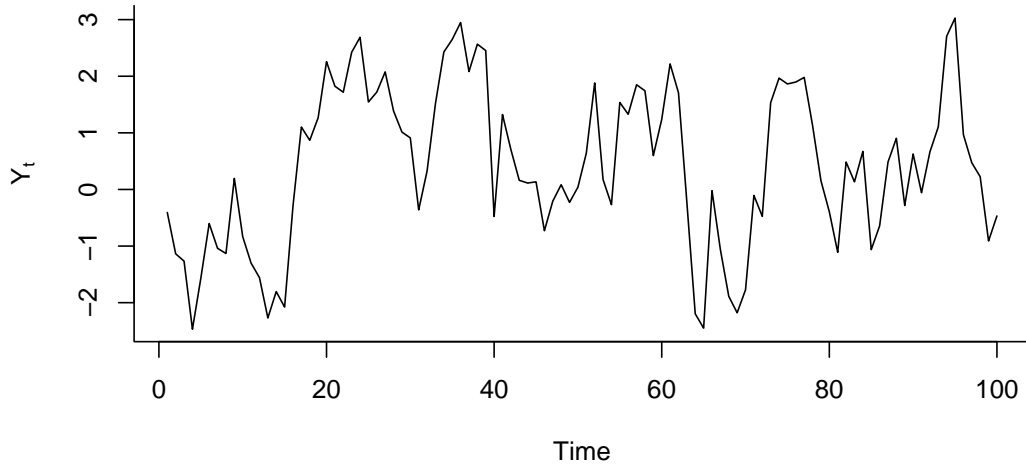


Figura 2: Processo  $AR(1)$  com  $\theta = 0.8$ .

A condição de estacionaridade para o processo  $AR(1)$  é então dado por  $|\theta| < 1$ . O caso importante de  $\theta = 1$  será abordado mais adiante. Um processo autoregressivo de ordem  $p$  pode então ser descrito como

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} + \varepsilon_t$$

onde,

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2).$$

A condição de estacionaridade para o processo generalizado  $AR(p)$  é que o inverso de todas as raízes do polinômio estejam dentro do círculo unitário. Um processo  $AR(1)$  estacionário também pode ser reescrito como um processo  $MA(\infty)$ . Isto pode ser feito de maneira mais fácil quando utilizamos um operador de defasagem,  $\mathbf{L}$ .

#### 4.3.1 O operador de defasagem, $\mathbf{L}$

Vejamos agora algumas propriedades básicas do operador de defasagem e como usá-lo.

**Propriedades básicas:**

1.  $L\alpha = \alpha$
2.  $Lx_t = x_{t-1}$

$$3. L^2 x_t = L\{Lx_t\} = Lx_{t-1} = x_{t-2}$$

$$4. (1 - L)x_t = x_t - Lx_t = x_t - x_{t-1} = \Delta x_t$$

$$5. L(1 - L)x_t = (1 - L)x_{t-1} = \Delta x_{t-1}$$

Agora, utilizando o operador de defasagem, podemos mostrar que um processo  $AR(1)$  pode ser reescrito como um processo  $MA(\infty)$ . Primeiro, definimos o processo  $AR(1)$  como

$$x_t = \alpha x_{t-1} + \varepsilon_t$$

onde  $|\alpha| < 1$ . Usando as propriedades (1) e (2) do operador de defasagem temos que

$$x_t = \alpha Lx_t + \varepsilon_t \implies (1 - \alpha L)x_t = \varepsilon_t$$

$$x_t = \frac{\varepsilon_t}{1 - \alpha L}$$

O último passo só será possível se  $(1 - \alpha L)$  for inversível. Dado que assumimos que  $|\alpha| < 1$ , sabemos que este processo é estacionário. Agora, lembrando novamente das propriedades de uma série geométrica, podemos reescrever esta última expressão como

$$x_t = \sum_{k=0}^{\infty} \alpha^k L^k \varepsilon_t$$

ou simplesmente

$$x_t = \sum_{k=0}^{\infty} \alpha^k \varepsilon_{t-k}$$

o que nada mais é do que um processo de médias móveis infinito.

### 4.3.2 Passeio aleatório

Suponha agora um processo  $AR(1)$  onde  $\theta = 1$  e  $\mu = 0$ , o que temos agora será um passeio aleatório dado por

$$Y_t = Y_{t-1} + \varepsilon_t$$

Que tal vermos se este processo é estacionário ou não?

$$\begin{aligned} E\{Y_T\} &= E\{Y_{t-1} + \varepsilon_t\} = E\{Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t\} = \dots \\ &= E\{Y_0 + \sum_{i=1}^T \varepsilon_i\} = Y_0 \end{aligned}$$

Como podemos ver, a esperança é constante e independe do tempo. Vejamos então a variância

$$\begin{aligned} V\{Y_T\} &= E\{(Y_t - E\{Y_t\})^2\} = E\{(Y_{t-1} + \varepsilon_{t-1} - Y_0)^2\} = \dots \\ &= E\{(Y_0 + \sum_{i=1}^T \varepsilon_i - Y_0)^2\} = \sum_{i=1}^T \sigma^2 = T\sigma^2 \end{aligned}$$

A variância é portanto uma função crescente do tempo, logo o passeio aleatório não é um processo estacionário. Note desde já que a persistência de um choque neste modelo é infinita. Por isso muitas vezes é chamado de um modelo com memória infinita.

**Raiz unitária:** Quando temos que  $\theta = 1$  aparece o problema de raiz unitária que, basicamente, é sinônimo de não estacionaridade. O nome surge do fato da solução da equação característica ter uma raiz igual a 1.

Um passeio aleatório pode se tornar estacionário ao ser diferenciado, isto é

$$Y_t = Y_{t-1} + \varepsilon_t \longrightarrow \Delta Y_t = \varepsilon_t.$$

Este fato será amplamente usado ao longo deste texto. Processos como este, que se tornam estacionários ao serem diferenciados, são chamados de estacionários em diferença.

**Processo estocástico integrado:** Um processo estocástico que possa ser feito estacionário através da diferenciação também é chamado de um processo estocástico integrado. Se um processo se torna estacionário ao se tomar sua primeira diferença dizemos que este processo é integrado de ordem um, ou  $I(1)$ . Da mesma maneira, se o processo somente se tornar estacionário na  $d$ -ésima diferença dizemos que este é um processo integrado de ordem  $d$ , ou  $I(d)$ . Um processo que não precisa ser diferenciado para se tornar estacionário é integrado de ordem zero, ou  $I(0)$ .

### 4.3.3 Passeio aleatório com *drift*

Usaremos agora o mesmo processo autoregressivo  $A(1)$  que antes com uma pequena mudança, onde assumimos que  $\mu \neq 0$ . O resultado é um modelo de passeio aleatório com *drift* dado por

$$Y_t = \mu + Y_{t-1} + \varepsilon_t$$

**Tendência estocástica:** A série  $Y_t$  descrita acima irá ter uma tendência positiva ou negativa dependendo do valor do parâmetro do *drift*  $\mu$ . Este *drift* também é chamado

de tendência estocástica. Vejamos agora que nem a esperança e nem a variância de um passeio aleatório são independentes do tempo.

$$\begin{aligned} E\{Y_T\} &= E\{\mu + Y_{T-1} + \varepsilon_T\} = E\{\mu + \mu + Y_{T-2} + \varepsilon_{T-1} + \varepsilon_T\} = \dots \\ &= E\left\{Y_0 + \sum_1^T \mu + \sum_1^T \varepsilon_t\right\} \\ &= Y_0 + T\mu \end{aligned}$$

Logo a esperança de  $Y_t$  depende do tempo. A variância deste processo pode ser calculada como

$$\begin{aligned} V\{Y_T\} &= E\{(Y_T - E\{Y_t\})^2\} = E\{(Y_T - Y_0 + T\mu)^2\} = E\{(\mu + Y_{T-1} + \varepsilon_T - Y_0 + T\mu)^2\} = \dots \\ &= E\left\{\sum_1^T \varepsilon_t^2\right\} = T\sigma_\varepsilon^2 \end{aligned}$$

Assim como com o passeio aleatório anterior, podemos tornar o passeio aleatório com *drift* estacionário ao tomar a primeira diferença

$$Y_t - Y_{t-1} = \mu + \varepsilon_t$$

$$\Delta Y_t = \mu + \varepsilon_t$$

Abaixo simulamos e visualizamos o passeio aleatório com *drift* e sua primeira diferença.

```
TT <- 100
et <- rnorm(TT)
mu <- 0.1
Yt <- ts(cumsum(mu+et))
dYt <- diff(Yt)
```

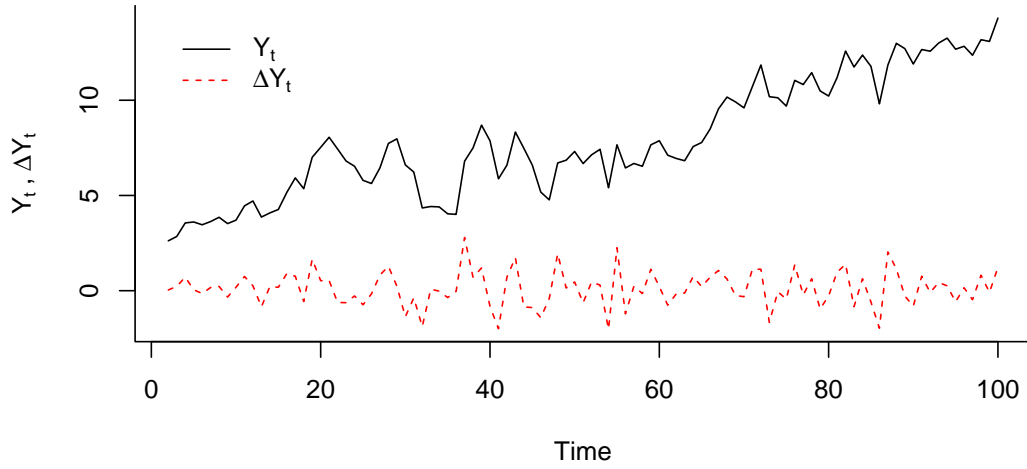


Figura 3: Passeio aleatório com drift e sua primeira diferença.

#### 4.3.4 Processos tendência-estacionários

Mais uma vez tomemos o processo autoregressivo  $AR(1)$  só que desta vez iremos assumir que  $\mu \neq 0$ ,  $|\theta| < 1$  e vamos adicionar uma tendência temporal,  $t$ . Assim teremos que

$$Y_t = \mu + t + \theta Y_{t-1} + \varepsilon_t$$

Este processo também será não estacionário com esperança dada por

$$E\{Y_t\} = \frac{\mu + t}{1 - \theta}$$

e variância igual a

$$V\{Y_t\} = \frac{\sigma_\varepsilon^2}{1 - \theta^2}$$

Podemos tornar esta série estacionária ao remover, ou subtrair, a tendência determinística<sup>8</sup> da série original. Processos como este, que podem ser tornados estacionários pela remoção de uma tendência determinística são chamados de processos tendência-estacionários.

---

<sup>8</sup>Note que tendências  $H-P$ , quadrática, cúbica e etc. são tendências determinísticas, apesar de não serem lineares.

## 5 Função de autocorrelação

Assumindo estacionaridade fraca, podemos definir a  $k$ -ésima ordem de autocovariância,  $\gamma_k$ , como

$$\gamma_k = cov\{Y_t, Y_{t-k}\} = cov\{Y_t, Y_{t+k}\}$$

A autocovariância de um processo estocástico pode ser normalizada e apresentada como uma função de autocorrelação,  $\rho_k$

$$\rho_k = \frac{cov\{Y_t, Y_{t-k}\}}{V\{Y_t\}} = \frac{\gamma_k}{\gamma_0}$$

A função de autocorrelação, ou comumente chamade de *ACF* do inglês, nos ajuda a caracterizar o desenvolvimento de  $Y_t$  ao longo do tempo. Ela nos mostra o quão forte o valor observado hoje está correlacionado com os valores observados no passado e como choques hoje afetam valores futuros da variável estocástica. Além de nos ajudar a descrever os dados, a *ACF* também nos ajuda a achar raízes unitárias, escolher modelos e fazer diagnósticos sobre regressões.

A *ACF* de um processo  $AR(1)$  é dado por

$$\rho_k = \frac{cov\{Y_t, Y_{t-k}\}}{V\{Y_t\}} = \frac{\theta^k \frac{\sigma_\varepsilon^2}{1-\theta^2}}{\frac{\sigma_\varepsilon^2}{1-\theta^2}} = \theta^k.$$

Para o processo  $MA(1)$  a *ACF* é

$$\rho_1 = \frac{cov\{Y_t, Y_{t-1}\}}{V\{Y_t\}} = \frac{\alpha \sigma_\varepsilon^2}{(1 + \alpha^2) \sigma_\varepsilon^2} = \frac{\alpha}{1 + \alpha^2}$$

porém,

$$\rho_k = 0 \quad , k > 1.$$

No **R** o comando `acf` calcula e nos dá um gráfico da *ACF* de uma série. Abaixo um exemplo de como simular e calcular a *ACF* é dado e, em seguida, os gráficos de 4 processos são dados.

```
TT <- 200

ar1 <- acf(arima.sim(model=list(ar=0.9), n=TT),
           ylab=expression(Y[t]),
           main=expression(paste("AR(1) com ", theta, "=0.9")))
```

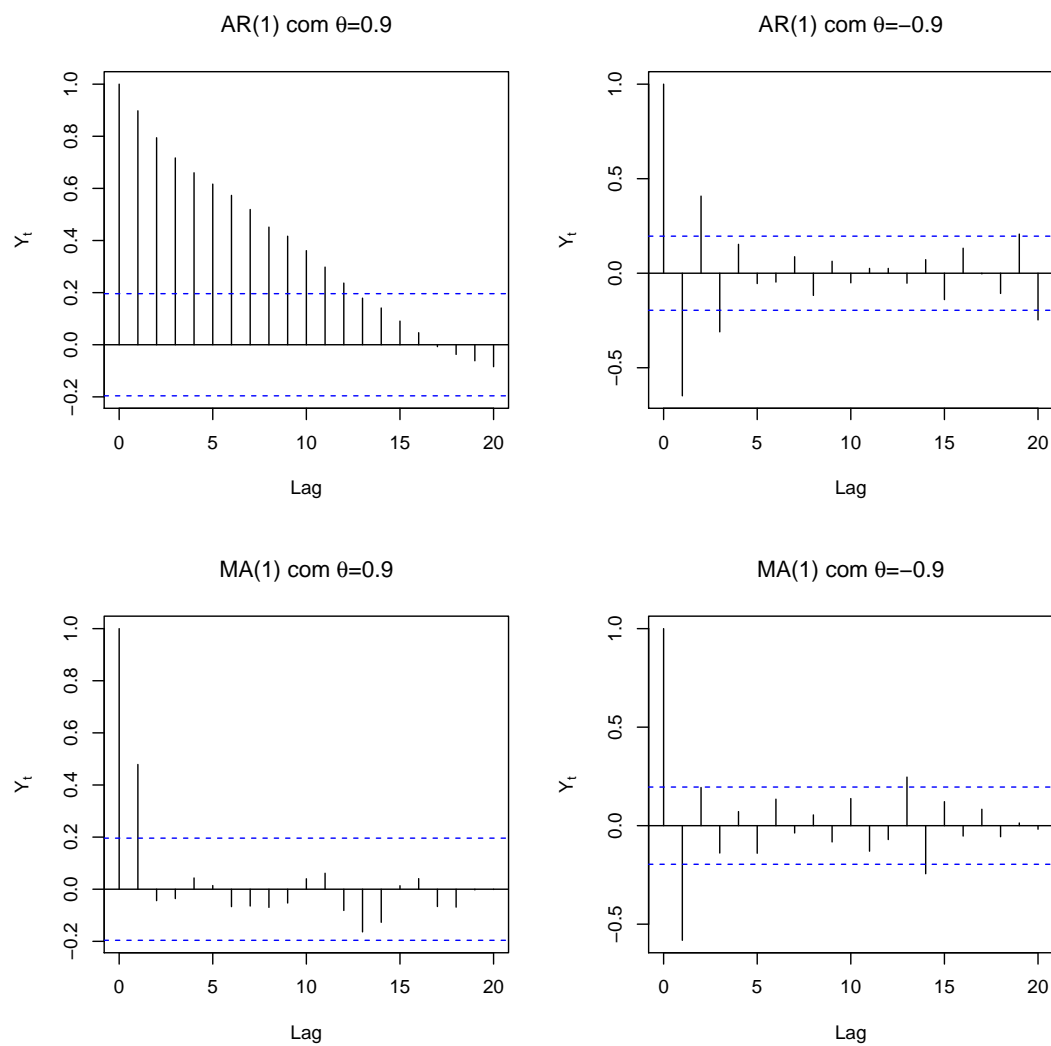


Figura 4: Função de autocorrelação para processos autoregressivos e de médias móveis.

## 6 Função de autocorrelação parcial

A função de autocorrelação parcial (ou comumente *PACF*, do inglês) nos dá correlação entre a variável no instante  $t$  e uma de suas defasagens, retirado os efeitos das outras defasagens. Por exemplo, assuma o seguinte processo  $AR(2)$

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_t.$$

O parâmetro  $\beta_2$  nos dá então o coeficiente de correlação entre  $Y_t$  e  $Y_{t-2}$  retirado os efeitos de  $Y_{t-1}$ . Assim,  $\beta_2$  pode ser descrito como a seguir

$$\beta_2 = r_{13,2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}}.$$

Assumindo estacionaridade fraca temos que

$$r_{12} = \text{corr}\{Y_t, Y_{t-1}\} = \text{corr}\{Y_{t-1}, Y_{t-2}\} = \rho_1$$

e

$$r_{13} = \text{corr}\{Y_t, Y_{t-2}\} = \rho_2.$$

Logo, concluímos que

$$\beta_2 = r_{13,2} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.$$

Assim como para as autocorrelações, o **R** possui um comando para calcular e plotar a autocorrelação parcial. Um exemplo é dado a seguir.

```
TT <- 200

ar1 <- pacf(arima.sim(model=list(ar=0.9), n=TT),
            ylab=expression(Y[t]),
            main=expression(paste("AR(1) com ", theta, "=0.9")))
```



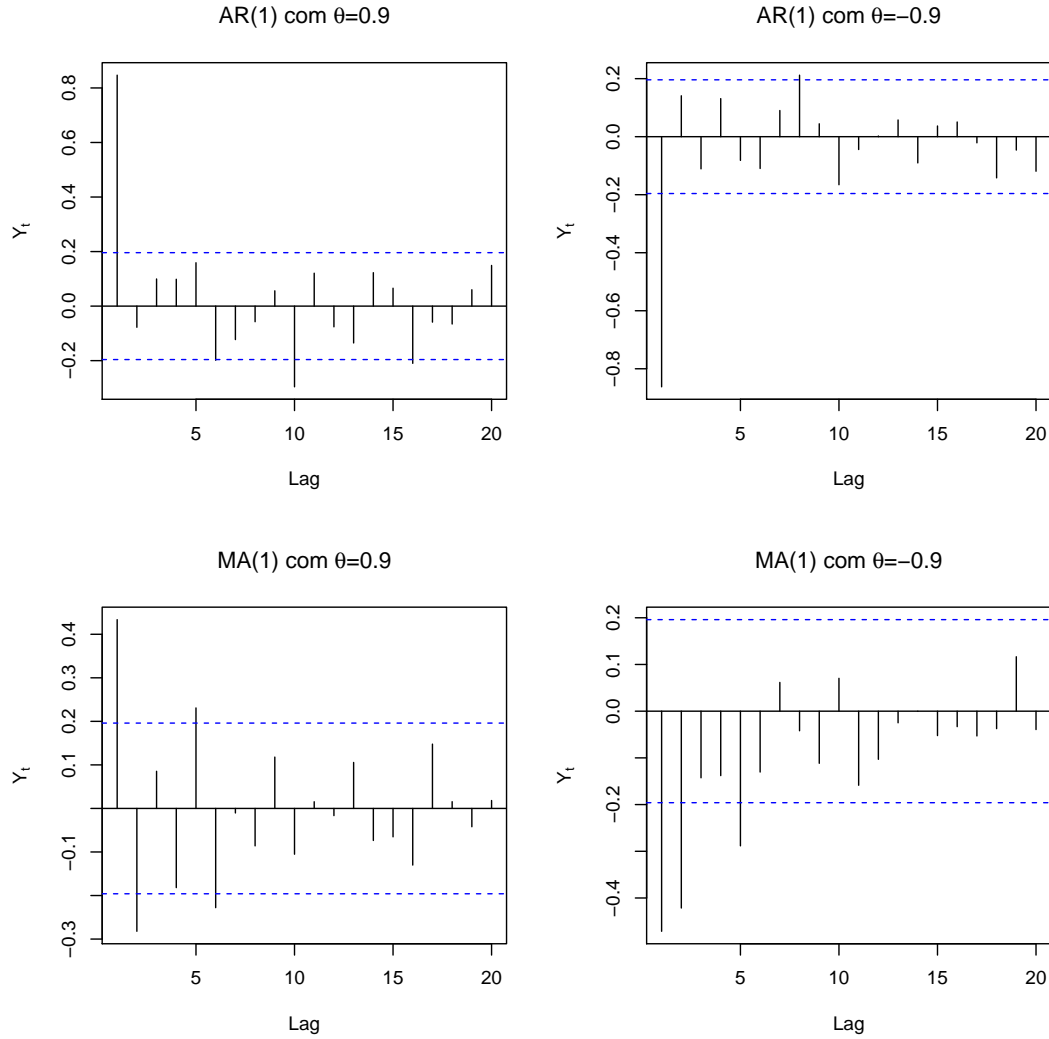


Figura 5: Função de autocorrelação parcial para processos autoregressivos e de médias móveis.

## 6.1 Um breve resumo dos resultados

Após estas duas seções podemos concluir que um processo  $AR(p)$  é descrito por

1. Uma  $ACF$  infinita e com queda exponencial;
2. uma  $PACF$  (aproximadamente) igual a zero para defasagens maiores que  $p$ .

Já o processo  $MA(q)$  é caracterizado por

1. Uma  $ACF$  (aproximadamente) igual a zero para defasagens maiores que  $q$ ;
2. uma  $PACF$  infinita e com queda exponencial;

Mais uma vez deve-se notar que esta "estratégia" de identificação do processo somente será válido quando se trata de uma série estacionária. Logo, antes de se começar a identificação do processo de uma série, devemos remover padrões como sazonalidade e tendências da série. Somente após esta etapa é que podemos analisar, de maneira correta, a *ACF* e *PACF*.

## 7 Processos ARMA

Agora que temos uma noção do que são processos autoregressivos e de média móvel podemos juntar os dois, desta forma podemos analisar processos mistos, ou processos *ARMA*. Este processo,  $ARMA(p, q)$ , mais geral pode ser formulado como a seguir

$$Y_t = \mu + \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q}$$

onde

$$\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2).$$

Para uma série não estacionária devemos formular um modelo  $ARIMA(p, d, q)$ , onde  $d$  nos dá o número de diferenças necessárias para tornar a série estacionária.<sup>9</sup>

### 7.1 Estimando modelos *ARIMA* - Intro

O objetivo da análise *ARIMA* é representar o processo dos resíduos de maneira parcimoniosa. Logo, devemos usando somente a quantidade necessária de termos *AR* e *MA*.

#### 7.1.1 A metodologia *Box-Jenkins* - Intro

A metodologia *Box-Jenkins* pressupõe que a série sendo analisado é estacionária, logo devemos tratar da possível existência de raiz unitária e sazonalidades. A metodologia pode ser brevemente caracterizada como a seguir.

##### 1. Identificação

- (a) Tratar da sazonalidade
- (b) Determinar a order de integração,  $d$
- (c) Achar valores apropriados para  $p$  e  $q$

##### 2. Estimação

---

<sup>9</sup>Nesta apostila não abordaremos modelos com sazonalidade, denotados por *SARIMA*.

- (a) Modelos puramente autoregressivos podem ser estimados de maneira consistente<sup>10</sup> por *OLS*, *OLS* não linear e por máxima verossimilhança.
- (b) Modelos puramente caracterizados por médias móveis podem ser estimados de maneira consistente por *OLS* não linear e por máxima verossimilhança.
- (c) Modelos *ARMA* podem ser estimados de maneira consistente por *OLS* não linear e por máxima verossimilhança.

### 3. Diagnóstico

- (a) Especificação do modelo
- (b) Autocorrelações
- (c) ARCH
- (d) Normalidade

---

<sup>10</sup>Consistência é uma propriedade assintótica, porém não garante estimadores não viesados em pequenas amostras.

## 8 Testes de estacionaridade

Como vimos anteriormente, o segundo passo da metodologia de *Box-Jenkins* é o de determinar a ordem de integração,  $d$ , de uma série<sup>11</sup>. Apresentamos a seguir alguns dos testes que existem para se determinar a ordem de integração de uma série.

### 8.1 O teste de *Dickey-Fuller*

Considere a seguir uma série caracterizada pelo seguinte processo

$$Y_t = \theta Y_{t-1} + \varepsilon_t$$

onde

$$\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2).$$

Se  $|\theta| < 1$  então teremos que o processo é estacionário. Se  $|\theta| = 1$  o processo apresentará uma raiz unitária e, portanto, será não estacionário. Logo, o teste de raiz unitária testa se  $\theta = 1$  ou não. Na prática fazemos primeiramente a seguinte transformação

$$Y_t - Y_{t-1} = \theta Y_{t-1} - Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = (\theta - 1)Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = \pi Y_{t-1} + \varepsilon_t.$$

Feitas estas transformações testamos a hipótese nula de que  $\pi = 0$ . Desta forma, se  $\pi = 0$  então  $\theta = 1$  e, conseqüentemente,  $Y_t$  possui raiz unitária. Note que a hipótese nula deste teste é dada por  $H_0 : \pi = 0$  enquanto que a alternativa é  $H_1 : \pi < 0$ , logo trata-se de um teste unilateral<sup>12</sup>.

**Exemplo:** *Determinando a ordem de integração - o teste DF*

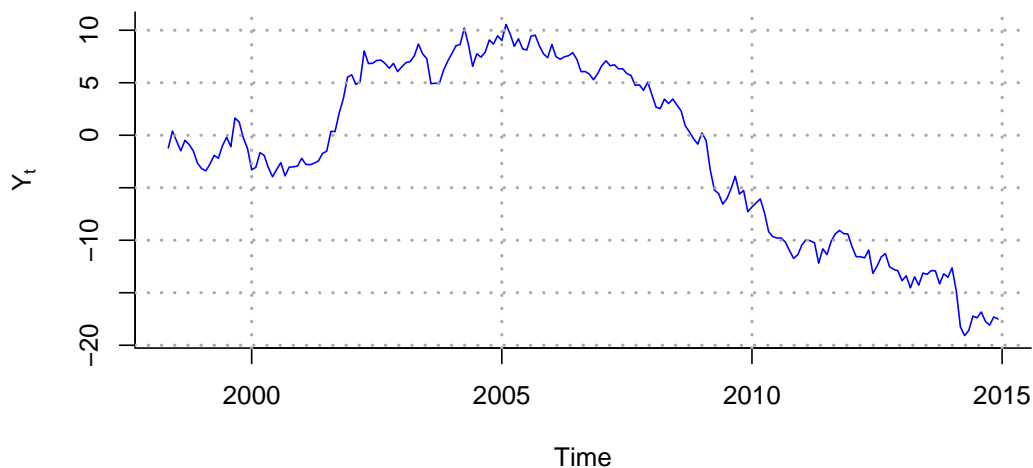
---

<sup>11</sup>Aqui abordamos somente métodos simples para remover sazonalidade dos dados e, para o nosso objetivo, assumiremos que estes serão suficientes.

<sup>12</sup>Note que sob a  $H_0$  o  $t$ -valor do coeficiente  $\hat{\pi}$  não segue uma distribuição  $t$ . Isto significa que a tabela  $t$  que comumente utilizamos para testes de hipótese não deve ser utilizada.

Assuma agora que temos em mãos uma certa série,  $y_t$ , e que queremos determinar sua ordem de integração, ou seja, quantas vezes nós devemos diferenciá-la de modo a torná-la estacionária. Primeiro, devemos sempre começar visualizando a série em questão.

```
plot(yt,
     ylab=expression(Y[t]),
     main='',
     bty='l',
     col='blue')
grid(col='darkgrey', lwd=2)
```



A série aparentemente é estacionária? Em geral, concluir se uma série é estacionária, ou não, simplesmente olhando para o seu gráfico é basicamente impossível. Para testar esta hipótese de maneira mais formal podemos então seguir as instruções dadas acima. Ou seja, rodamos a regressão da primeira diferença da série contra sua defasagem.<sup>13</sup>

```
summary(lm(diff(yt)~lag(yt,-1)[-length(yt)] - 1))

##
```

---

<sup>13</sup>Note que no código desta passagem adicionamos o  $-1$ , este comando força a regressão através da origem, ou seja, sem intercepto. Além disso temos que corrigir o tamanho do vetor das observações da defasagem.

```
## Call:
## lm(formula = diff(yt) ~ lag(yt, -1)[-length(yt)] - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3856 -0.7056 -0.0622  0.6434  2.8990
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## lag(yt, -1)[-length(yt)] 0.003974    0.008372   0.475    0.636
##
## Residual standard error: 0.9889 on 198 degrees of freedom
## Multiple R-squared:  0.001137, Adjusted R-squared:  -0.003908
## F-statistic: 0.2253 on 1 and 198 DF,  p-value: 0.6355
```

Como podemos notar, a estatística  $t$  para a hipótese nula de  $\pi = 0$  é igual a 0.47, logo não rejeitaríamos a hipótese nula ao seguir a distribuição padrão  $t$ . Porém, para testar de maneira correta esta hipótese, devemos nos guiar pela tabela desenvolvida por Dickey and Fuller (1979), esta é dada por exemplo em Enders (2009). Para o nosso caso temos que o valor crítico é aproximadamente  $-2.58$ , logo não podemos rejeitar a hipótese nula. Assim, concluímos que a série  $y_t$  é não estacionária<sup>14</sup>.

Como veremos agora, o uso de um pacote do **R** pode facilitar bastante a nossa vida. Apesar de podermos especificar manualmente as regressões de interesse (e isto contribuir para criar uma intuição), o trabalho fica muito mais prático e rápido se tivermos acesso a processos já automatizados. Assim, de agora e em diante utilizaremos o pacote **urca**. Para isso devemos carregar o pacote<sup>15</sup>. Abaixo carregamos o pacote necessário e, em seguida, rodamos um teste  $DF$ .

---

<sup>14</sup>Note que como o teste é unilateral temos que  $|\tau| > \tau_{\text{crítico}}$  não necessariamente significa que rejeitamos a hipótese nula. Isto pode ser verificado se lembrarmos que a estatística de teste só será positiva se valer que  $\theta > 1$ , algo que implica na não estacionaridade da série.

<sup>15</sup>Uma preferência pessoal que tenho é a de sempre começar um script do **R** declarando e carregando os pacotes necessários para todos os meus comandos.

```

library(urca)
summary(ur.df(yt, type='none', lags=0))

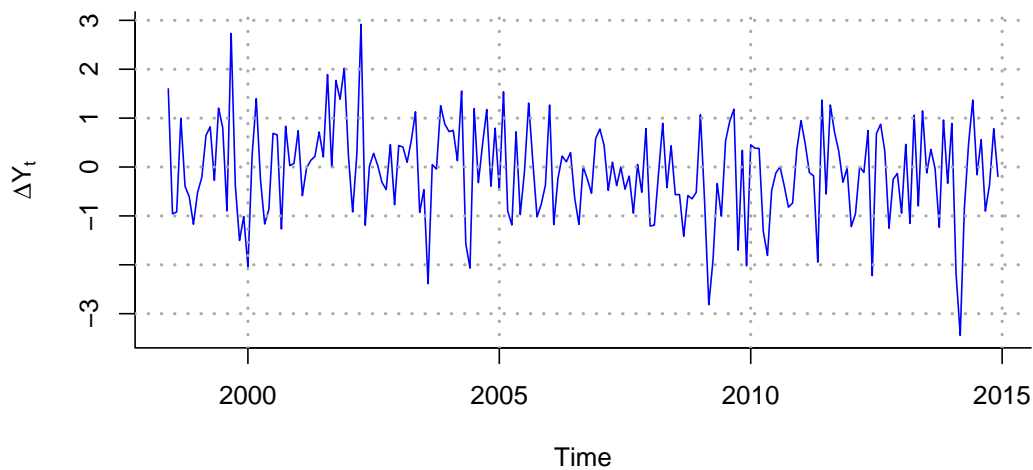
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3856 -0.7056 -0.0622  0.6434  2.8990
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## z.lag.1  0.003974    0.008372   0.475    0.636
##
## Residual standard error: 0.9889 on 198 degrees of freedom
## Multiple R-squared:  0.001137, Adjusted R-squared:  -0.003908
## F-statistic: 0.2253 on 1 and 198 DF,  p-value: 0.6355
##
##
## Value of test-statistic is: 0.4747
##
## Critical values for test statistics:
##      1pct   5pct 10pct
## tau1 -2.58 -1.95 -1.62

```



Como podemos ver, a função `ur.df` nos retorna automaticamente os valores críticos para o nosso teste, logo não precisamos ter a tabela por perto. Agora a única que resta fazer é comparar o valor de  $t$  com o valor crítico  $\tau$ . Se  $t < \tau$  rejeitamos a hipótese nula,  $H_0 : \pi = 0$ , ou seja, rejeitamos que nossa série possui raiz unitária.

Neste nosso caso específico, como  $t > \tau$ , não rejeitamos a hipótese de que a série possui raiz unitária. Logo, devemos prosseguir com os nossos testes. Agora devemos testar para ver se nossa série é  $I(1)$ , pois acabamos de concluir que ela não é  $I(0)$ . Para isso diferenciamos a nossa série, visualizamos o gráfico dela e, finalmente, fazemos o teste  $DF$  em sua primeira diferença.



```
dym <- diff(yt)
summary(ur.df(dym, type='none', lags=0))

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
```

```
## lm(formula = z.diff ~ z.lag.1 - 1)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -3.3479 -0.6935 -0.0649  0.5995  2.9076
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1 -0.95564      0.07071  -13.52  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9844 on 197 degrees of freedom
## Multiple R-squared:  0.4811, Adjusted R-squared:  0.4785
## F-statistic: 182.6 on 1 and 197 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -13.5146
##
## Critical values for test statistics:
##      1pct   5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

Desta vez rejeitamos a hipótese de que  $\pi = 0$ , i.e. de raiz unitária, e, consequentemente, podemos concluir que segundo o teste de *Dickey-Fuller* a série  $y_t$  é  $I(1)$ . Ou seja,

$$\text{Se } t < \tau \implies \text{Rejeita-se } H_0 : \pi = 0$$

Desta vez podemos afirmar com certeza absoluta que o processo  $y_t$  é  $I(1)$ , pois fomos nós que criamos ele. Como pode ser visto abaixo no código que utilizamos, a série  $y_t$  é um passeio aleatório.

```
yt = ts(cumsum(rnorm(200)), end=c(2014,12), freq=12)
```

Este caso do teste  $DF$  é, obviamente, o mais simples de todos. É importante notar que diferentes versões deste teste significa o uso de diferentes distribuições de  $\tau$ . Estes diferentes valores também estão reportados em Enders (2009) e se referem a testes como

$$\Delta Y_t = \pi Y_{t-1} + \varepsilon_t,$$

$$\Delta Y_t = \alpha + \pi Y_{t-1} + \varepsilon_t,$$

$$\Delta Y_t = \alpha + \beta t + \pi Y_{t-1} + \varepsilon_t,$$

ou seja, testes  $DF$  que incorporam uma constante e também termos de tendência. Sob a hipótese nula a estatística  $F$  também se modificam e seus valores críticos  $\phi$  também podem ser encontrados em Enders (2009).

A não ser que haja forte razões teóricas, devemos sempre começar testando a presença de raiz unitária a partir de um modelo com constante e tendência.

## 8.2 O Poder do teste de *Dickey-Fuller* de raiz unitária

**Erro Tipo II:** A hipótese nula não é rejeitada quando a alternativa é verdadeira. Quando o valor verdadeiro do nosso parâmetro varia muito em relação a nossa hipótese nula (por exemplo, no caso de um processo  $AR(1)$ ,  $\theta = 0.01$  e nossa  $H_0$  é  $\theta = 1$ ) a probabilidade de cometermos um erro do tipo II é bem pequena. Do contrário, se estes forem próximos um do outro, a probabilidade de cometer um erro do tipo II será significativamente grande (por exemplo,  $\theta = 0.98$  e  $H_0$  de  $\theta = 1$ ).

**O Poder do Teste:** O poder do teste é 1 menos a probabilidade de se cometer um erro do tipo II, isto é, a probabilidade de se rejeitar a hipótese nula quando ela efetivamente é falsa<sup>16</sup>.

O poder dos testes  $DF$  e  $ADF$  é, em geral, pequeno devido a alta persistência e/ou tendência presente nas variáveis macroeconômicas. A consequência disto é que muitas vezes tendemos a aceitar a existência de uma raiz unitária quando na verdade deveríamos

---

<sup>16</sup>O "tamanho" do teste é a percentagem de vezes que rejeitamos uma hipótese nula que na verdade deveríamos aceitar.

rejeita-lá. Este problema, ao se incluir um número maior de regressores que não estão presentes no processo da série, somente aumenta<sup>17</sup>.

### 8.3 O teste de *Dickey-Fuller* Aumentado

O teste  $DF$  assume que os erros são não correlacionados, ou seja, que a covariância entre  $\varepsilon_t$  e  $\varepsilon_{t-k}$  para todo  $k > 0$  seja nula. Em se tratando de dados *cross-section* a existência de autocorrelação serial implica em estimadores ineficientes, porém não viesados. Já no caso de processos autoregressivos, comumente presentes em séries de tempo, esta correlação terá como consequência estimados ineficientes e viesados. Logo, neste caso, a inferência só pode ser feita após garantirmos que os resíduos não apresentam autocorrelação. Assim, após rodarmos o teste  $DF$  faz-se necessário verificar se os resíduos da regressão são autocorrelacionados, por exemplo através dos correlogramas.

Se esta hipótese não valer podemos aumentar o teste  $DF$  de forma a incluir um número maior de defasagens das diferenças. Isto é, podemos utilizar uma regressão do tipo

$$\Delta Y_t = \alpha + \beta t + \pi Y_{t-1} + \sum_{i=1}^m \gamma_i \Delta Y_{t-i} + \varepsilon_t.$$

O problema agora será o de determinar o número de defasagens utilizado no teste. Dado que o poder do teste é baixo, queremos utilizar o menor número possível de defasagens que garanta a não autocorrelação dos resíduos. Uma maneira de prosseguir seria o de começar com um número "exagerado" de defasagens e reduzir de acordo com um teste de correlação. Alternativamente, pode-se começar sem defasagens e adicionar até chegarmos a um número que resulta em rejeição da hipótese de autocorrelação. Existe porém um risco com este último método dado que defasagens posteriores possam ser significativas, logo estariam sendo omitidas do modelo<sup>18</sup>. Outros métodos incluem: reduzir o número de defasagens através de testes  $t$ , determinação por critérios de informação ( $AIC$ ,  $BIC$ , por exemplo).

---

<sup>17</sup>Ou seja, incluir regressores "desnecessários" somente agrava o problema.

<sup>18</sup>Este risco pode ser ainda maior em dados com sazonalidade.

## 8.4 O teste de *Phillips-Perron*

Como visto anteriormente, o teste *DF* assume que os resíduos são não correlacionados, sendo que uma maneira de garantir isso é a de incluir defasagens das diferenças da variável dependente na regressão de teste, isto é, aplicar o teste *ADF*.

O teste de *Phillips-Perron*, *PP*, lida com o problema de correlação dos resíduos de uma maneira diferente, sem a adição de defasagens<sup>19</sup>. O teste baseia-se na mesma equação do teste *DF* com constante e tendência, porém utiliza o estimador *Newey-West* para estimar a matriz de variância robusta<sup>20</sup>. O teste *PP*, utilizando a função `PP.test` usa a hipótese nula de  $\theta = 1$  porém tem como alternativa  $\theta < 1$ .

De forma a ilustrar o uso deste teste podemos simular duas séries, a primeira sendo um passeio aleatório e outra sendo um ruído branco, como a seguir.

```
set.seed(12)
t <- 200
y1 <- ts(cumsum(rnorm(t)))
y2 <- rnorm(t)
```

O gráfico das séries é dado abaixo, assim como o resultado dos testes.

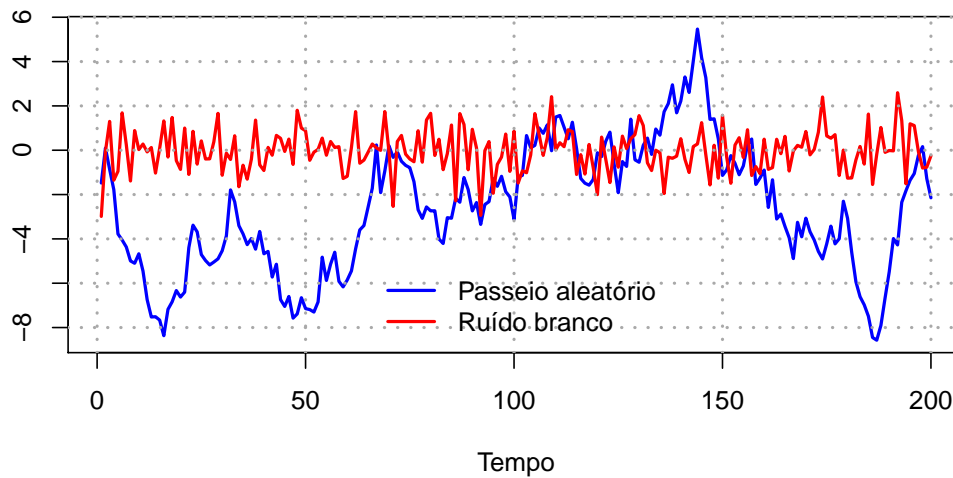
```
ts.plot(y1, y2,
        ylab='', xlab='Tempo',
        main='',
        col=c('blue', 'red'),
        lwd=2)
grid(col='darkgrey', lwd=2)

legend('bottom', c('Passeio aleatório', 'Ruído branco'),
       lwd=2, col=c('blue', 'red'), bty='n')
```

---

<sup>19</sup>A abordagem do teste *PP* é não paramétrica.

<sup>20</sup>Isto se for utilizada a função `PP.test` do **R**.



```
PP.test(y1)
```

Phillips-Perron Unit Root Test

data: y1 Dickey-Fuller = -2.6661, Truncation lag parameter = 4, p-value = 0.2971

```
PP.test(y2)
```

Phillips-Perron Unit Root Test

data: y2 Dickey-Fuller = -15.0281, Truncation lag parameter = 4, p-value = 0.01

Como podemos observar, no primeiro caso não rejeitamos a hipótese nula, isto é, a série apresenta raiz unitária. Enquanto que no segundo caso rejeitamos a hipótese nula, confirmando assim que o ruído branco não possui raiz unitária.

## 8.5 O teste *ADF* sequencial

Como acabamos de ver, devemos partir da especificação mais geral quando testamos a presença de raiz unitária, ou seja

$$\Delta Y_t = \alpha + \beta t + \pi Y_t + \sum_{i=1}^m \gamma_i \Delta Y_{t-i} + \varepsilon_t.$$

Além disso, sabemos que a inclusão de regressores desnecessários irá diminuir o poder do teste e, conseqüentemente, iremos tender a aceitar a hipótese nula  $H_0 : \pi = 0$ . Logo,

se mesmo utilizando esta especificação rejeitarmos a hipótese nula podemos concluir que a série  $Y_t$  não possui raiz unitária.

Porém, se o contrário acontecer, isto é, se não rejeitarmos a hipótese nula, devemos suspeitar do poder do teste. Portanto, o primeiro passo é o de testar a significância da tendência, ou seja, testar se  $\beta = 0$  dado que  $\pi = 0$  <sup>21</sup>.

Se não rejeitarmos esta hipótese ( $\beta = 0|\pi = 0$ ) podemos prosseguir e testar a hipótese de que  $\pi = 0$  sob normalidade, ou seja, utilizando a estatística  $t$ . Não rejeitando esta hipótese ( $\beta = 0|\pi = 0$ ) estamos concluindo que devemos incluir a tendência na regressão de teste e após testar se  $\pi = 0$  poderemos concluir se a série apresenta ou não raiz unitária.

Considere agora o caso em que a tendência seja não significativa, isto é,  $\beta = 0|\pi = 0$  não é rejeitado. Logo, devemos retirar a tendência e reestimar o teste. Feito isto, repetimos o processo acima; testamos a hipótese de que  $\pi = 0$  (agora usando a estatística de teste  $\tau_2$ )<sup>22</sup>, se rejeitarmos podemos concluir que a série possui raiz unitária, caso contrário, voltamos a suspeitar do poder do teste.

Não rejeitando voltamos a testar a significância do *drift*. Logo, testamos se  $\alpha = 0$  dado que  $\pi = 0$ <sup>23</sup>. Se rejeitarmos  $\alpha = 0|\pi = 0$  podemos prosseguir e testar  $\pi = 0$  sob normalidade. Caso contrário, reestimamos a regressão de teste sem o termo de *drift* e, finalmente, testar se  $\pi = 0$  utilizando a estatística de teste  $\tau_1$ .

Aceitando esta hipótese podemos concluir que a série  $Y_t$  é, no mínimo,  $I(1)$ . Porém, para determinar a ordem de integração da série será preciso repetir desde o início o processo descrito acima, só que desta vez utilizando a primeira diferença da série,  $\Delta Y_t$ .

Utilizando esta estratégia seremos capazes de determinar se uma série é:

1. Estacionária com média zero

---

<sup>21</sup>Segundo Pfaff (2008), devemos utilizar para este teste o valor crítico de  $\phi_3$ .

<sup>22</sup>Ainda seguindo Pfaff (2008).

<sup>23</sup>Para isto utilizamos a estatística de teste  $\phi_1$ .

2. Estacionária com média diferente de zero

3. Possui raiz unitária sem *drift*

4. Possui raiz unitária com *drift*

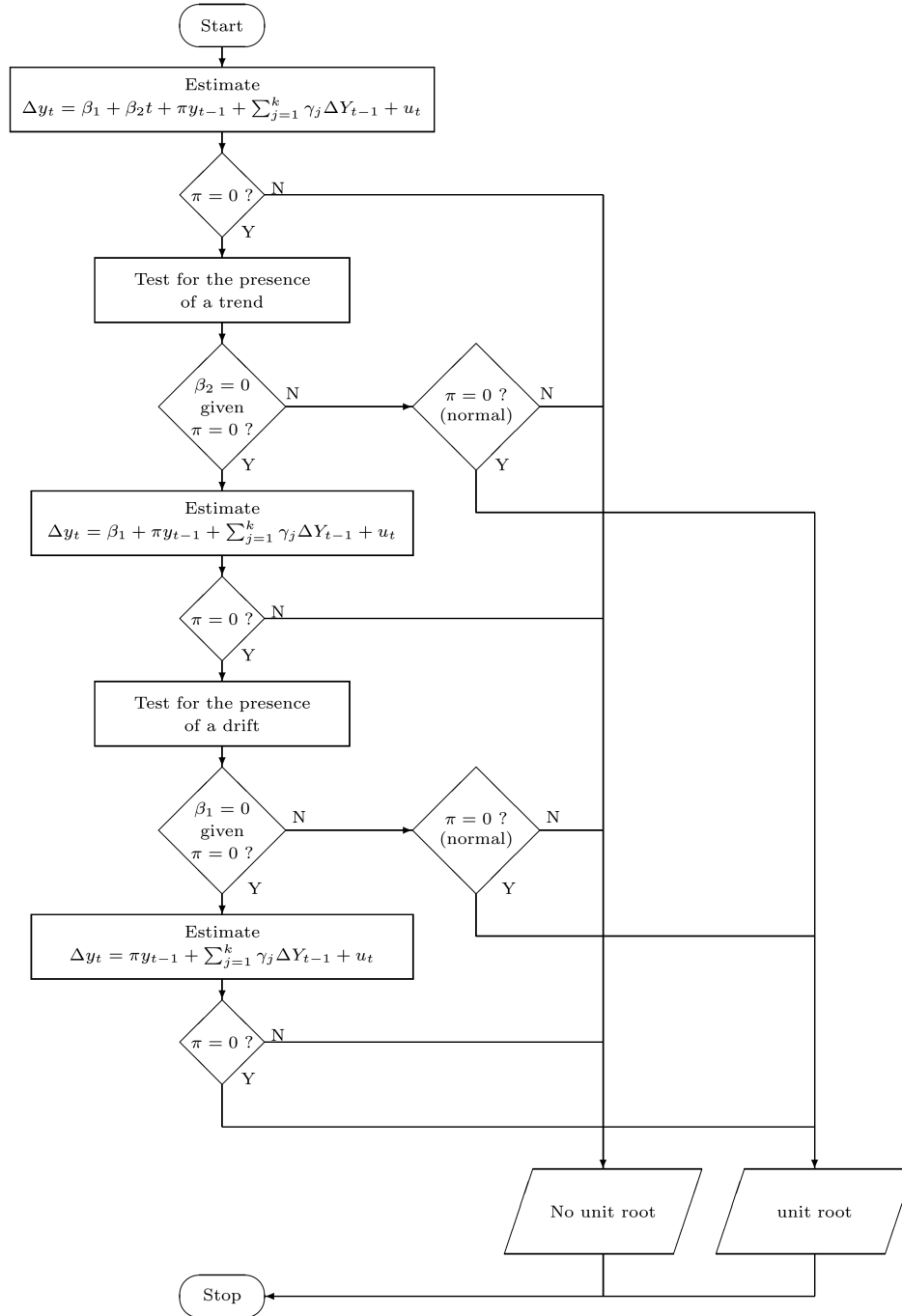


Figura 6: Fluxograma para teste de raiz unitária.



A Figura 6, retirada de Pfaff (2008)<sup>24</sup>, mostra o fluxograma de decisões, testes e conclusões que podemos fazer quando utilizamos esta estratégia<sup>25</sup>. Vejamos a seguir um exemplo dessa estratégia de determinação de ordem de integração.

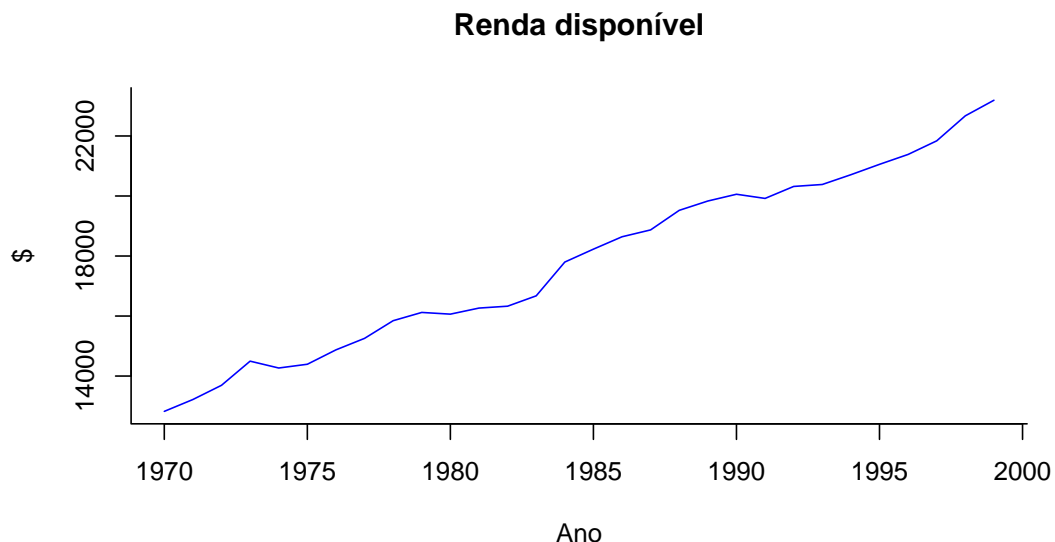
**Exemplo:** *Determinando a ordem de integração - o teste ADF sequencial*

Para este exemplo iremos utilizar o arquivo `table17-2.txt`, os dados são anuais e referentes a gastos com consumo e renda disponível. No código a seguir importamos os dados e indicamos para o **R** que trata-se de dados anuais com início em 1970. Como o nosso interesse é a renda disponível (PPDI) iremos extrair somente esta variável do arquivo.

```
y <- read.table('table17-2.txt', header=T)$PPDI  
  
y <- ts(y, start = 1970, freq=1)
```

Feita a importação podemos, como sempre, começar pela visualização dos dados.

```
plot(y, col='blue', main='Renda disponível',  
      bty='l', xlab='Ano', ylab='$')
```



---

<sup>24</sup>Uma versão parecida pode ser encontrada em Enders (2009).

<sup>25</sup>Note que todas as estatísticas de teste envolvidas ( $\tau_1, \tau_2, \tau_3, \phi_1$  e  $\phi_3$ ) são dadas, com a mesma conotação, pelo pacote `ur.df`.

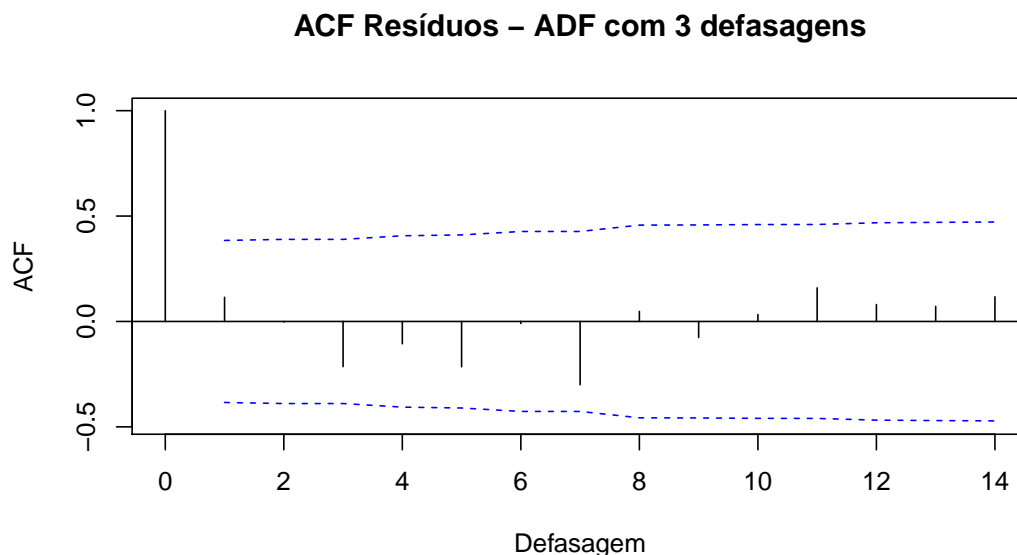
Como podemos ver, a série tem uma clara tendência positiva. Logo, devemos testar pela presença de um processo de raiz unitária. O primeiro passo será então estimar o modelo mais completo, isto é, com constante e tendência, e três<sup>26</sup> diferenças defasadas.

```
library(urca)

adf.y <- ur.df(y, type='trend', lags=3)
```

Estimada a regressão podemos analisar os resíduos que, como vimos anteriormente, devem ser não correlacionados para que o teste seja válido.

```
acf(adf.y@res, ci.type='ma', main='ACF Resíduos - ADF com 3 defasagens',
    xlab='Defasagem')
```



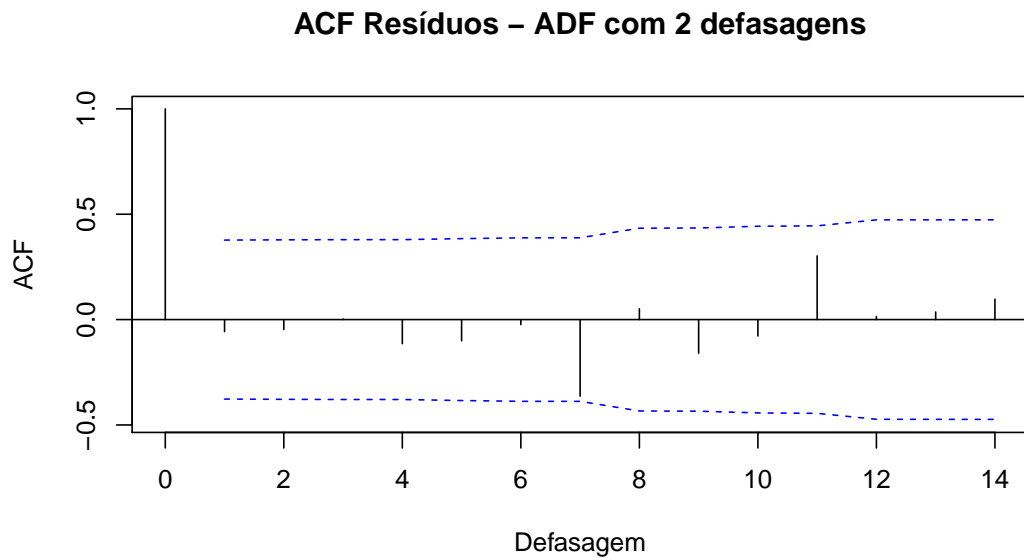
Como nenhuma autocorrelação é estatisticamente significativa talvez seja possível aumentar o poder do teste ao retirarmos uma defasagem do teste. Rodamos agora o teste *ADF* com somente duas defasagens e analisamos novamente o correlograma dos resíduos.

---

<sup>26</sup>Aqui utilizaremos a regra de bolso  $3 * \textit{frequencia}$  para determinar o número inicial de defasagens a serem incluídas na regressão.

```
adf.y <- ur.df(y, type='trend', lags=2)

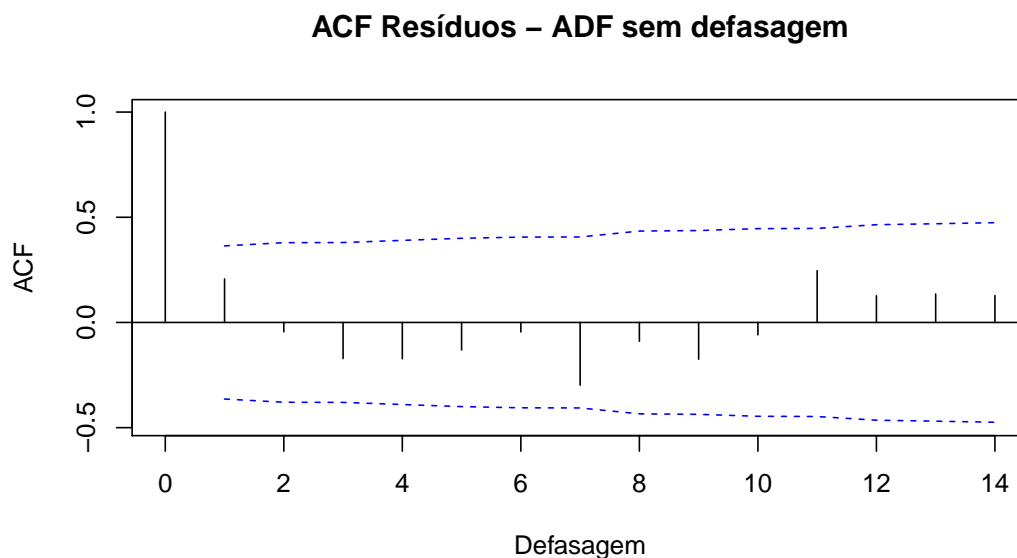
acf(adf.y$res, ci.type='ma', main='ACF Resíduos - ADF com 2 defasagens',
    xlab='Defasagem')
```



Mais uma vez nenhuma defasagem é estatisticamente significativa, podemos então repetir o procedimento anterior.

```
adf.y <- ur.df(y, type='trend', lags=0)

acf(adf.y$res, ci.type='ma', main='ACF Resíduos - ADF sem defasagem',
    xlab='Defasagem')
```



Como pode ser visto no gráfico anterior, continuando com o procedimento iremos concluir que não é necessário incluir nenhuma defasagem no teste. Agora podemos finalmente começar a testar nossas hipóteses. A primeira é que  $\pi = 0$  e para isso devemos comparar o valor  $t_{\pi}$  com estatística de teste  $\tau_3$ .

```
summary(adf.y)

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -417.27 -212.19   10.75  134.86  547.15
##
```

```

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4792.8567  2016.7164   2.377  0.0251 *
## z.lag.1      -0.3549    0.1597  -2.222  0.0352 *
## tt           121.6977    53.8866   2.258  0.0325 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 271.1 on 26 degrees of freedom
## Multiple R-squared:  0.165, Adjusted R-squared:  0.1008
## F-statistic:  2.57 on 2 and 26 DF,  p-value: 0.09586
##
##
## Value of test-statistic is: -2.2223 18.5193 2.5696
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -4.15 -3.50 -3.18
## phi2  7.02  5.13  4.31
## phi3  9.31  6.73  5.61

```

Dado que  $t_\pi > \tau_3$ , não podemos rejeitar a hipótese de que  $\pi = 0$ , portanto devemos procurar aumentar o poder do teste. Para isso podemos testar se a tendência é estatisticamente significativa com a hipótese  $\beta = 0 | \pi = 0$ . A estatística para esta hipótese é dada por  $\phi_3$  que, utilizando um nível de significância de 5%, é igual a 6.73. Como estamos querendo testar se todas os coeficientes são não significativos, isto é  $\beta = 0 | \pi = 0$ , a estatística de teste estimada é dada por  $F$ , que é igual a 2.57. Como  $F < \phi_3$ , não podemos rejeitar que  $\beta = 0 | \pi = 0$ , logo podemos concluir que o termo da tendência pode ser retirado para que o poder do teste fique maior.

Agora que a tendência determinística foi retirada devemos refazer estes procedimentos utilizando uma constante, ou o *drift*, na equação de teste.

```

adfd.y <- ur.df(y, type='drift', lags=0)

summary(adfd.y)

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -574.47 -140.23   -7.91  132.03  771.88
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.955e+02  3.420e+02   0.864   0.395
## z.lag.1      3.516e-03  1.914e-02   0.184   0.856
##
## Residual standard error: 291 on 27 degrees of freedom
## Multiple R-squared:  0.001248, Adjusted R-squared:  -0.03574
## F-statistic: 0.03373 on 1 and 27 DF,  p-value: 0.8557
##
##
## Value of test-statistic is: 0.1836 21.9026
##
## Critical values for test statistics:
##      1pct   5pct 10pct

```

```
## tau2 -3.58 -2.93 -2.60
## phi1  7.06  4.86  3.94
```

Aqui, novamente, não podemos rejeitar a hipótese de raiz unitária pois  $t_\pi = 0.18 > \tau_2 = -2.93$ . Para verificar se podemos aumentar o poder do teste testamos se  $\alpha = 0 | \pi = 0$ , comparando a estatística de teste<sup>27</sup> com seu valor crítico,  $\hat{\phi} = 21.90 > \phi_1 = 4.86$ . Portanto, podemos rejeitar que  $\alpha = 0 | \pi = 0$ , e concluimos que não devemos retirar o termo de *drift* da regressão de teste.

Agora, podemos testar se  $\pi = 0$  sob normalidade, isto é, comparando  $t_\pi$  com o  $t$  de Student. Obviamente, como  $t_\pi = 0.18 > 0$ , podemos concluir<sup>28</sup> rapidamente que trata-se de uma processo com raiz unitária.

Para confirmar que esta série é, de fato,  $I(1)$  devemos recomençar estes testes utilizando agora a primeira diferença como variável independente.

Completar com:

1. Teste de raiz unitária com quebra estrutural
2. Raiz unitária sazonal (HEGY)

---

<sup>27</sup>Note que agora, como se trata de um termo de intercepto, não utilizamos a estatística  $F$ .

<sup>28</sup>Lembre-se que  $\pi = \theta - 1$ , logo se estatística de teste  $t$  for maior que zero significa que  $\theta > 1$  e, consequentemente, o processo terá uma raiz unitária.

## 9 Estimando modelos ARIMA

Agora que já entendemos um pouco mais sobre os pré-requisitos do processo de estimação, isto é, que devemos analisar graficamente os dados, "limpá-los" e "filtra-los", podemos prosseguir para a estimação de um modelo **ARIMA**.

Já deve estar bem claro agora que o principal objetivo da análise econométrica de séries de tempo é de caracterizar um fenômeno da maneira mais parcimoniosa possível. Devemos utilizar um número suficiente de termos **AR** e **MA** de forma a garantir que os resíduos sejam completamente aleatórios, ou seja, que sejam caracterizados por um ruído branco.

Para este objetivo, iremos utilizar a metodologia apresentada em Dickey and Fuller (1979), que citamos, de maneira breve, anteriormente.

### 9.1 A metodologia *Box-Jenkins*

Relembrando, temos que a metodologia *Box-Jenkins* pode ser brevemente caracterizada pelos seguintes "passos":

#### 1. Identificação

- (a) Tratar da sazonalidade
- (b) Determinar a ordem de integração,  $d$
- (c) Achar valores apropriados para  $p$  e  $q$

#### 2. Estimação

- (a) Modelos puramente autoregressivos podem ser estimados de maneira consistente<sup>29</sup> por *OLS*, *OLS* não linear e por máxima verossimilhança.
- (b) Modelos puramente caracterizados por médias móveis podem ser estimados de maneira consistente por *OLS* não linear e por máxima verossimilhança.

---

<sup>29</sup>Consistência é uma propriedade assintótica, porém não garante estimadores não viesados em pequenas amostras.



- (c) Modelos *ARMA* podem ser estimados de maneira consistente por *OLS* não linear e por máxima verossimilhança.

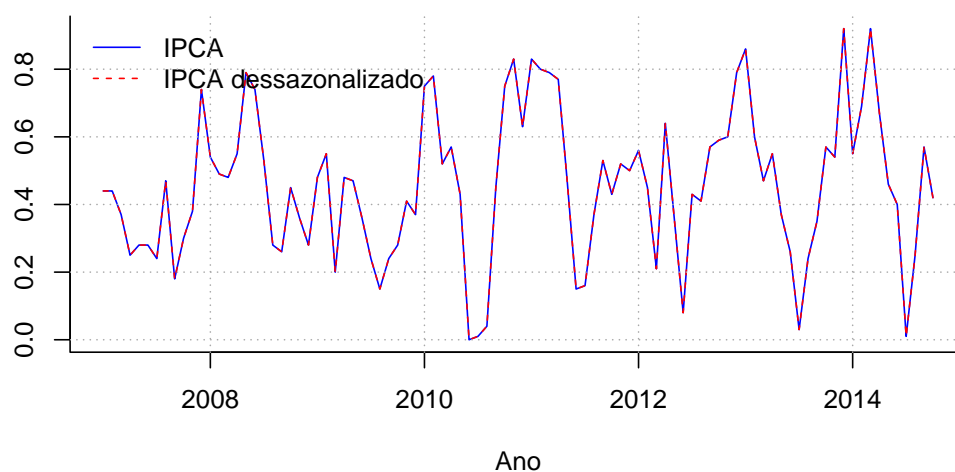
### 3. Diagnóstico

- (a) Especificação do modelo
- (b) Autocorrelações
- (c) ARCH
- (d) Normalidade

Relembrado a base da estrutura dessa metodologia que tal um exemplo prático ?

## 9.2 A metodologia *Box-Jenkins* - Um exemplo prático

Para este exemplo iremos utilizar a série mensal da variação mensal do IPCA retirada do *SIDRA*. A série tem seu início em 1980 porém iremos utilizar a série a partir de 2007. Aqui, de modo a simplificar nossa análise, trataremos da sazonalidade de maneira determinística. Isto é, seguiremos os procedimentos feitos anteriormente na apostila. A seguir temos o gráfico da série original e sua versão "filtrada", isto é, a variação mensal do IPCA retirado os efeitos de sazonalidade.



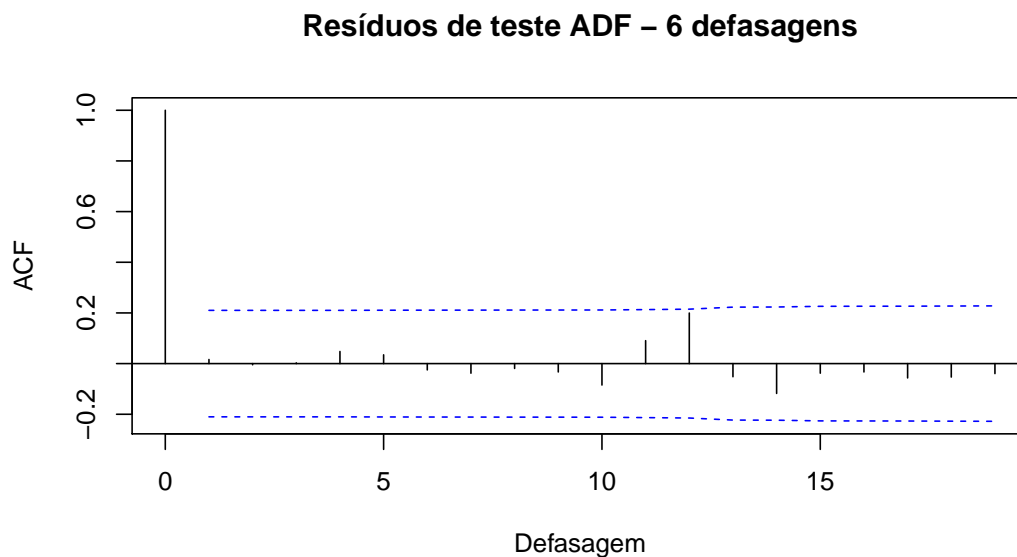
Como pode ser visto, a retirada da sazonalidade determinística teve basicamente zero efeito sobre nossa série. Logo, as *dummies* de meses provavelmente são estatisticamente iguais a zero. Agora, vamos tentar identificar a ordem de integração da série.

### 9.2.1 Ordem de integração

Para isso, começamos pelo teste *ADF* com constante e 36 defasagens<sup>30</sup>, verificando o *ACF* dos resíduos da regressão.

```
adf1 <- ur.df(ipcad, type='drift', lags=6)

acf(adf1$res, ci.type='ma',
    main='Resíduos de teste ADF - 6 defasagens',
    xlab='Defasagem')
```



Começando pela regressão de teste com 36 defasagens, podemos concluir, com o uso dos correlogramas, que podemos aumentar o poder do teste usando somente 6 termos de defasagem. A seguir, temos o resultado do teste *ADF* incluindo 6 defasagens.

```
summary(adf1)

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
```

---

<sup>30</sup>Relembrando a regra de bolso de 3\*frequencia.

```
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37502 -0.10580 -0.01029  0.12128  0.34183
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.47659    0.08553   5.572 3.37e-07 ***
## z.lag.1       -1.02606    0.18080  -5.675 2.20e-07 ***
## z.diff.lag1    0.57071    0.15813   3.609 0.000537 ***
## z.diff.lag2    0.42860    0.14605   2.935 0.004370 **
## z.diff.lag3    0.47473    0.13354   3.555 0.000642 ***
## z.diff.lag4    0.32352    0.13263   2.439 0.016957 *
## z.diff.lag5    0.24056    0.12129   1.983 0.050814 .
## z.diff.lag6    0.23950    0.11403   2.100 0.038884 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1725 on 79 degrees of freedom
## Multiple R-squared:  0.348, Adjusted R-squared:  0.2903
## F-statistic: 6.024 on 7 and 79 DF,  p-value: 1.212e-05
##
##
## Value of test-statistic is: -5.6751 16.1136
##
## Critical values for test statistics:
```

```
##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

Logo, primeiro testamos a  $H_0 : \pi = 0$ . Como  $-4.85 < -2.89$ , podemos rejeitar a existência de uma raiz unitária. Ou seja, podemos concluir que a variação mensal do IPCA é  $I(0)$  segundo o teste *ADF*.<sup>31</sup>

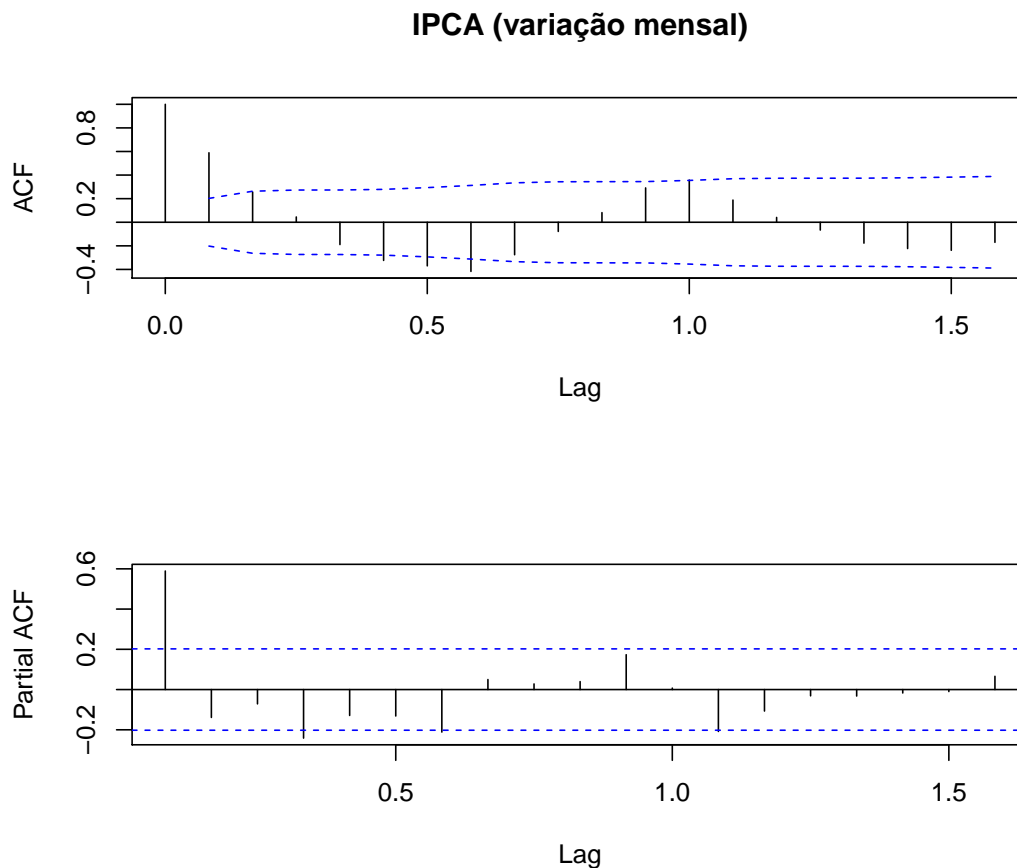
### 9.2.2 Escolhendo $p$ e $q$

Agora que identificamos que a variação mensal do IPCA é estacionária, queremos caracterizar esta parte estacionária do processo. Faremos isso utilizando o *ACF* e *PACF* de modo a identificar a ordem de termos **ARMA** que devemos utilizar. Abaixo temos os dois correlogramas.

```
par(mfrow=c(2,1))
acf(ipcad, ci.type='ma', main='IPCA (variação mensal)')
pacf(ipcad, main='')
```

---

<sup>31</sup>Talvez esta série tenha sido um exemplo muito fácil para aprendermos o intuito do teste *ADF* sequencial. Poderíamos talvez tentar a série trimestral do PIB?



Olhando para os correlogramas temos, o que pode parecer, um decaimento exponencial em ambos gráficos (bem de leve) sendo que a primeira defasagem do  $ACF$  é bastante significativa. Apesar de ser bem difícil identificar a partir dos correlogramas um modelo  $ARMA$  a ser estimado, é bem possível se tratar de algo como um  $ARMA(1,0)$ . Podemos começar com algo mais geral, como por exemplo um  $ARMA(2,2)$  e retirar termos não significantes.

### 9.2.3 Estimando um modelo ARMA

Prosseguiremos agora com a estimação de um modelo  $ARMA(2,2)$ .

```
coeftest(arima(ipcad, order=c(2,0,2)))

##
## z test of coefficients:
##
##          Estimate Std. Error z value Pr(>|z|)
```

```
## ar1      0.029864    0.987519    0.0302    0.9759
## ar2      0.205342    0.448281    0.4581    0.6469
## ma1      0.632196    0.983157    0.6430    0.5202
## ma2      0.061693    0.260710    0.2366    0.8129
## intercept 0.454590    0.039221  11.5905    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como podemos ver, nenhum coeficiente é estatisticamente diferente de zero, exceto a constante. Podemos começar reestimando o modelo retirando a defasagem de maior ordem e maior  $p$ -valor, que neste caso é o coeficiente *ma2*.

```
coeftest(arima(ipcad, order=c(2,0,1)))

##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1      0.287620    1.078856    0.2666    0.7898
## ar2      0.093555    0.670168    0.1396    0.8890
## ma1      0.374399    1.060180    0.3531    0.7240
## intercept 0.454544    0.039339  11.5546    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Continuando com a mesma estratégia, retiramos o coeficiente *ar2* e reestimamos o modelo.

```
coeftest(arima(ipcad, order=c(1,0,1)))

##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
```

```
## ar1          0.44348      0.15543   2.8532   0.004328 **
## ma1          0.21866      0.17357   1.2598   0.207745
## intercept    0.45451      0.03881  11.7111 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente, retiramos o último termo de média móvel e estimamos o modelo final.

```
coeftest(arima(ipcad, order=c(1,0,0)))

##
## z test of coefficients:
##
##          Estimate Std. Error z value Pr(>|z|)
## ar1          0.583238    0.082428   7.0757 1.487e-12 ***
## intercept    0.454794    0.042716  10.6470 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assim, nosso modelo que descreve o comportamento da variação mensal do IPCA é dado por

$$IPCA_t = 0.45 + 0.58IPCA_{t-1} + \varepsilon_t$$

#### 9.2.4 Diagnósticos dos resíduos

Agora que temos um modelo final para a inflação mensal medida pelo IPCA devemos ver se realmente captamos toda a estrutura do processo, isto é, devemos checar se os nossos resíduos estão "limpos". Devemos analisar as seguintes questões<sup>32</sup>:

1. Os resíduos da regressão são não autocorrelacionados?
2. Os resíduos são normalmente distribuídos?

---

<sup>32</sup>Note que estas questões que devemos analisar estão diretamente ligadas às hipóteses que fizemos para garantir a validade do nosso modelo. Estamos basicamente tentando garantir que podemos fazer inferência a partir da nossa estimativa e que o modelo seja válido para observações fora da amostra.

### 3. Devemos incluir/excluir regressores ?

#### Os resíduos da regressão são não autocorrelacionados?

Para responder a primeira questão podemos utilizar o teste *Ljung-Box*. Este teste tem como hipótese nula a independência de uma dada série de tempo, isto é, assume que as observações que temos (neste caso os resíduos de uma regressão) são conjuntamente não correlacionados ao longo do tempo. A não independência dos resíduos seria um indicador de que não capturamos alguma estrutura fora do modelo e, por isso, estaríamos errando de maneira sistematica. Este teste pode ser feito utilizando a função `Box.test`, com a opção de `type='Ljung-Box'`. A seguir testamos a hipótese nula de que os resíduos são independentes até a sexta defasagem.

```
Box.test(resid(reg.ipca), lag=6, type='Ljung-Box', fitdf=1)
```

Box-Ljung test

data: resid(reg.ipca) X-squared = 6.5934, df = 5, p-value = 0.2527

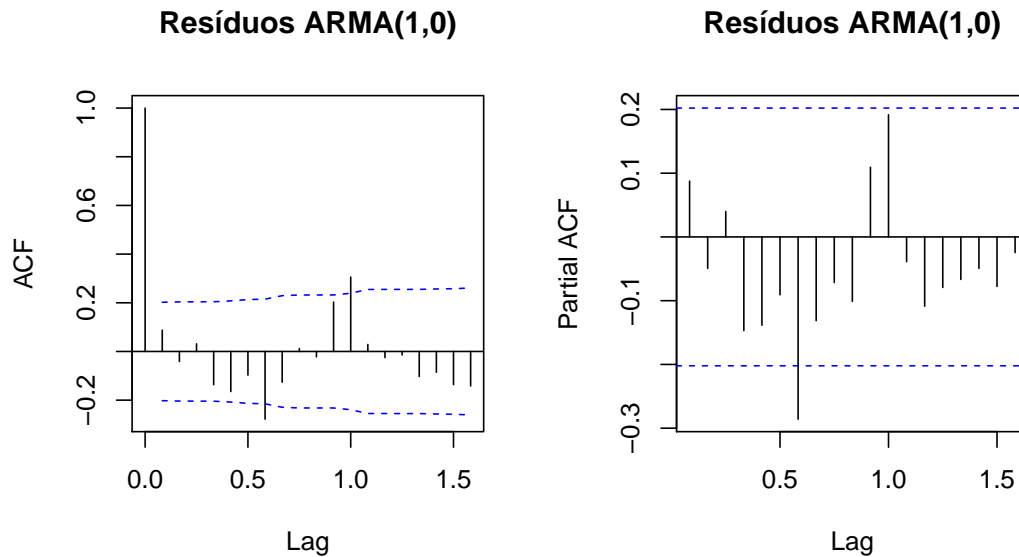
Note que usamos `fitdf=1`<sup>33</sup>, isto é feito para corrigir o grau de liberdade do teste dado que a série testada é resultado de uma regressão com 1 termo *AR*. Como podemos ver a estatística de teste é igual a 6.6 com *p*-valor de 0.25, logo não rejeitamos a hipótese nula e podemos concluir que os resíduos são conjuntamente não correlacionados (pelo menos até a sexta defasagem). A seguir os correlogramas dos resíduos são expostos.

```
par(mfrow=c(1,2))
acf(resid(reg.ipca), main='Resíduos ARMA(1,0)', ci.type='ma')
pacf(resid(reg.ipca), main='Resíduos ARMA(1,0)')
```

---

<sup>33</sup>Onde `fitdf` é "fit degrees of freedom"





Agora podemos ver as potenciais fraquezas do uso do teste *Ljung-Box*. Primeiro, temos que o *Ljung-Box* testa a significância conjunta das defasagens de uma série, logo não verifica a significância individual. Em segundo lugar, temos que o teste obviamente só irá testar até a ordem que por nós for especificada. Como pode ser visto pelo gráfico anterior, devemos suspeitar de algum padrão após a sexta defasagem. De modo a verificar essa nossa suspeita podemos realizar outro teste *Ljung-Box* só que desta vez testando até a sétima defasagem.

```
Box.test(resid(reg.ipca), lag=7, type='Ljung-Box', fitdf=1)
```

Box-Ljung test

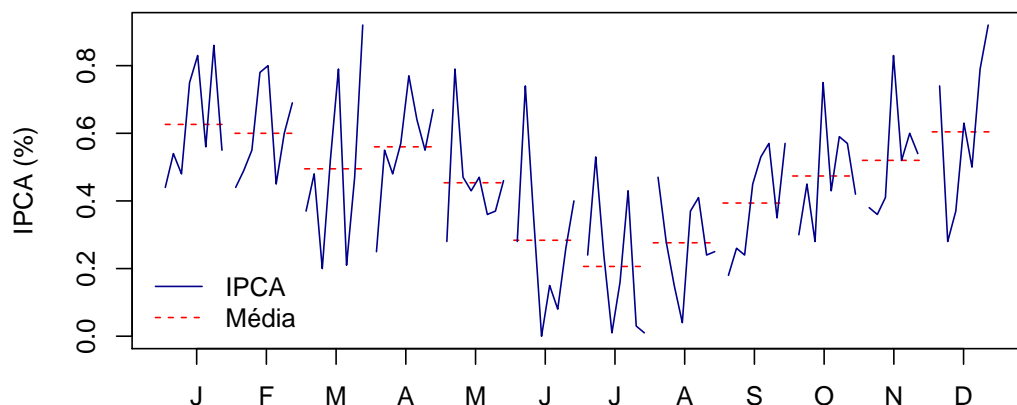
data: resid(reg.ipca) X-squared = 14.624, df = 6, p-value = 0.02339

Como podemos ver, nossa suspeita não foi em vão. Quais poderiam ser os possíveis problemas aqui ? Os mais interessados devem buscar a resposta, que na verdade não é tão difícil assim. Abaixo vai uma dica:

```
monthplot(ipca, col='darkblue', col.base='red',
           main='IPCA por mês e IPCA médio por mês',
           ylab='IPCA (%)', lty.base=2)

legend('bottomleft', legend=c('IPCA', 'Média'), lty=c(1,2),
       col=c('darkblue', 'red'), bty='n')
```

### IPCA por mês e IPCA médio por mês



### Extra: Criando uma função

Dado essa "fraqueza" no teste, talvez deveríamos fazer o teste *Ljung-Box* desde a primeira defasagem até a  $n$ -ésima defasagem, desta forma seríamos capazes de saber exatamente a partir de qual momento que começamos a rejeitar ou aceitar a hipótese. São neste momentos que o **R** se mostra bastante eficiente. O problema que queremos resolver é: como rodar o teste *Ljung-Box* utilizando a opção `lag` indo de 1 até  $n$ . Que tal criarmos uma função?

```
ljung.n <- function(dados, lag.final, fitdf=1){  
  
  # Pré-allocando um vetor que irá receber os p-valores  
  # dos testes Ljung-Box  
  
  pvalores <- matrix(NA, nrow=lag.final, ncol=1)  
  
  # Para cada valor entre 1 e o número máximo de  
  # defasagens a ser testada  
  
  for (i in 1:lag.final){
```

```

# Rodar teste Ljung-Box e guardar p-valor no vetor
# 'pvalores'

pvalores[i] <- Box.test(dados, fitdf=fitdf, lag=i)$p.value

}

# Retornar o vetor com os resultados

return(pvalores)
}

```

### Os resíduos são normalmente distribuídos?

O teste que iremos utilizar para responder esta pergunta é o teste de normalidade de *Shapiro-Wilk*, onde a hipótese nula é de que a série analisada é normalmente distribuída. Abaixo o resultado deste teste é apresentado.

```

shapiro.test(resid(reg.ipca))

##
##  Shapiro-Wilk normality test
##
## data:  resid(reg.ipca)
## W = 0.9911, p-value = 0.7874

```

Como o  $p$ -valor é de 0.79, não podemos rejeitar que os resíduos desta regressão são normalmente distribuídos.

### Devemos incluir/excluir regressores ?

Finalmente, para concluir esta seção de diagnósticos do nosso modelo proposto para o IPCA mensal, devemos comparar o modelo final com modelos alternativos. Seria possível explicar mais da variação do IPCA utilizando mais/menos regressores? Para esta tarefa podemos utilizar critérios de informação como o de *Akaike* ou o *Bayesiano*. Estes

critérios são construídos com base nos resíduos das regressões e representam, colocando de uma maneira bem simples, médias ponderadas. Logo, quanto menor for o critério de informação, melhor será o poder de explicação do modelo.

De modo a manter esta análise curta, iremos somente utilizar o critério de informação de *Akaike*, ou *AIC*. Como o nosso modelo final é um  $ARMA(1,0)$ , que tal comparar este com um  $ARMA(0,1)$ ,  $ARMA(1,1)$  e  $ARMA(2,2)$ ? Abaixo, estimamos estes modelos e calculamos os seus respectivos *AIC*s.

```
AIC(reg.ipca)                                # ARMA(1,0)

## [1] -54.34407

AIC(update(reg.ipca, order=c(0,0,1)))        # ARMA(0,1)

## [1] -49.84835

AIC(update(reg.ipca, order=c(1,0,1)))        # ARMA(1,1)

## [1] -53.97261

AIC(update(reg.ipca, order=c(2,0,2)))        # ARMA(2,2)

## [1] -50.0395
```

Note que aqui utilizamos a função `update` que nos permite "atualizar" um objeto que era resultado de uma função, algo que facilita bastante nossas operações<sup>34</sup>. Além disso, podemos concluir que nosso modelo  $ARMA(1,0)$  é de fato o "melhor" modelo, pelo menos de acordo com o *AIC*.

---

<sup>34</sup>Talvez aqui também seria útil montar uma função?

## 10 Previsão com modelos ARIMA

Agora que temos um modelo, o nosso interesse passa a ser o uso dele para fazer previsões sobre o comportamento futuro do IPCA. Antes de prosseguir com a previsão, vejamos alguns aspectos mais técnicos sobre o processo de previsão.

### 10.1 Usando modelos *ARIMA* para construir previsões

Dado o conjunto de informação  $I_T = \{Y_{-\infty}, \dots, Y_T\}$ , queremos escolher o preditor ótimo,  $Y_{T+h|I_T}$ , que minimize o erro quadrático médio esperado, ou seja, o problema que queremos resolver é dado por

$$\min E \left\{ \left( Y_{T+h} - \hat{Y}_{T+h|T} \right)^2 \mid I_T \right\}.$$

#### 10.1.1 O caso do processo autoregressivo

Tome como exemplo um processo *AR* simples

$$Y_t = \theta Y_{t-1} + \varepsilon_t$$

Assumindo este processo, temos que  $Y_{T+1} = \theta Y_T + \varepsilon_{T+1}$  e, conseqüentemente, o preditor ótimo para um passo a frente será

$$\begin{aligned} Y_{T+1|T} &= E \{ Y_{T+1} \mid Y_T, \dots, Y_{T-\infty} \} \\ &= E \{ (\theta Y_T + \varepsilon_{T+1}) \mid Y_T, \dots, Y_{T-\infty} \} \\ &= \theta Y_T + E \{ \varepsilon_{T+1} \mid Y_T, \dots, Y_{T-\infty} \} \\ Y_{T+1|T} &= \theta Y_T. \end{aligned}$$

De maneira parecida podemos derivar o preditor ótimo para dois passos a frente.

$$\begin{aligned} Y_{T+2|T} &= E \{ Y_{T+2} \mid Y_T, \dots, Y_{T-\infty} \} \\ &= E \{ (\theta Y_{T+1} + \varepsilon_{T+2}) \mid Y_T, \dots, Y_{T-\infty} \} \end{aligned}$$

$$= \theta E \{ \theta Y_T + \varepsilon_{T+1} | Y_T, \dots, Y_{T-\infty} \}$$

$$Y_{T+2|T} = \theta^2 Y_T.$$

Assim, generalizando temos que o preditor ótimo de um processo  $AR(1)$  para  $h$  passos à frente é dado por

$$Y_{T+h|T} = \theta^h Y_T.$$

Caso o processo  $AR(1)$  não seja de média zero, ou seja,  $Y_t = \mu + \theta Y_{t-1} + \varepsilon_t$ , teremos que o preditor ótimo será

$$Y_{T+h|T} = \mu + \theta^h (Y_T - \mu)$$

### 10.1.2 O caso do processo de médias móveis

Assuma agora um processo de médias móveis simples, representado por

$$Y_t = \varepsilon_t + \alpha \varepsilon_{t-1}.$$

Logo, temos que  $Y_{T+1} = \varepsilon_{T+1} + \alpha \varepsilon_T$  e o preditor um passo à frente será dado por

$$Y_{T+1|T} = E \{ Y_{T+1} | Y_T, \dots, Y_{T-\infty} \}$$

$$= E \{ \varepsilon_{T+1} + \alpha \varepsilon_T | Y_T, \dots, Y_{T-\infty} \}$$

$$= \alpha E \{ \varepsilon_T | Y_T, \dots, Y_{T-\infty} \}$$

$$Y_{T+1|T} = \alpha \varepsilon_T.$$

Para o caso de dois passos à frente teremos

$$Y_{T+2|T} = E \{ Y_{T+2} | Y_T, \dots, Y_{T-\infty} \}$$

$$= E \{ \varepsilon_{T+2} + \alpha \varepsilon_{T+1} | Y_T, \dots, Y_{T-\infty} \}$$

$$= \alpha E \{ \varepsilon_{T+1} | Y_T, \dots, Y_{T-\infty} \}$$

$$Y_{T+2|T} = 0.$$

Logo, diferentemente do caso autoregressivo, nossa melhor previsão para  $h$  passos à frente, quando  $h$  for maior que a ordem do processo de média móvel, será igual a zero.

## 10.2 Prevendo com o R

Dada esta pequena introdução teórica, podemos prosseguir e começar a fazer previsões. Para isso, iremos utilizar o modelo estimado anteriormente para o IPCA. Utilizando a função `predict` vamos fazer uma previsão de 4 meses para o IPCA.

```
prev <- predict(reg.ipca, n.ahead=4)
```

Como podemos ver, a função `predict` nos dá como resultado as estimativas e seus respectivos desvios-padrão. Podemos utilizar estas séries para fazer um gráfico mostrando nossas previsões. A seguir duas maneiras de fazer os gráficos são expostas, uma mais simples e outra um pouco mais complicada.

```
par(mfrow=c(1,2))

ts.plot(window(ipca, start=2013),
         prev$pred,
         prev$pred+1.96*prev$se,
         prev$pred-1.96*prev$se,
         col=c(1,2,2,2), lty=c(1,1,2,2))

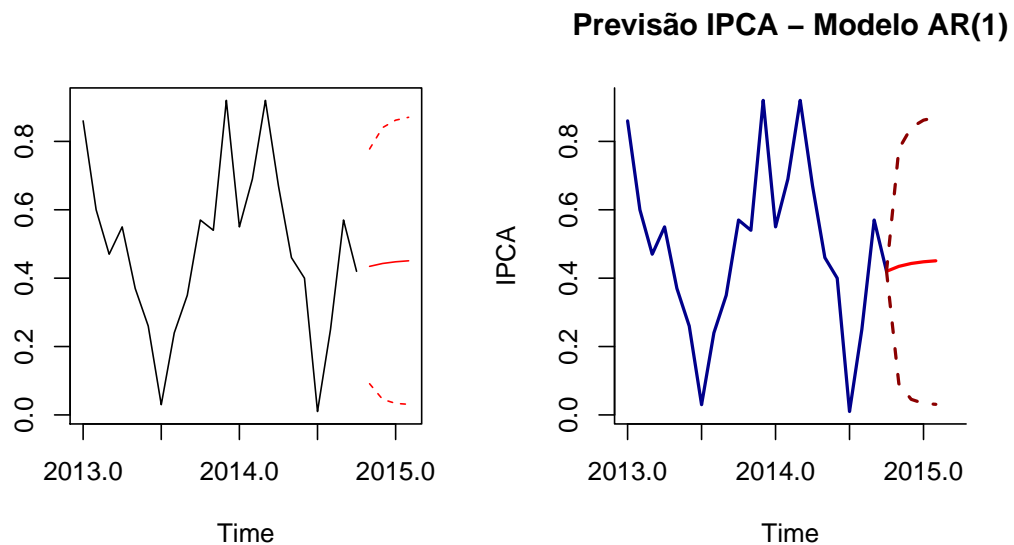
plot(window(ipca, start=2013),
      xlim=c(2013, 2015.2),
      col='darkblue', lwd=2, bty='l', ylab='IPCA',
      main='Previsão IPCA - Modelo AR(1)')

lines(ts(c(ipca[length(ipca)]),prev$pred), start=c(2014,10), freq=12),
      col='red', lwd=2)

lines(ts(c(ipca[length(ipca)]),prev$pred+1.96*prev$se),
      start=c(2014,10), freq=12),
      col='darkred', lwd=2, lty=2)

lines(ts(c(ipca[length(ipca)]),prev$pred-1.96*prev$se),
```

```
start=c(2014,10), freq=12),
col='darkred', lwd=2, lty=2)
```



Agora, como exercício, iremos remover as 4 últimas observações da nossa amostra do IPCA, re-estimar o modelo e, em seguida, fazer a previsão para 4 meses à frente e comparar com os valores observados. Como podem notar, este método irá fazer previsões para valores dentro da nossa amostra, enquanto que no primeiro caso fizemos previsões para valores fora da amostra.

```
ipca.b <- window(ipca, end=c(2014, 06))

reg.ipca.b <- arima(ipca.b, order=c(1,0,0))

prev2 <- predict(reg.ipca.b, n.ahead=4)
```

A seguir mostramos graficamente os resultados desse exercício.

```
par(mfrow=c(1,2))

ts.plot(window(ipca, start=2013),
        prev2$pred,
        prev2$pred+1.96*prev2$se,
        prev2$pred-1.96*prev2$se,
```



```

col=c(1,2,2,2), lty=c(1,1,2,2))

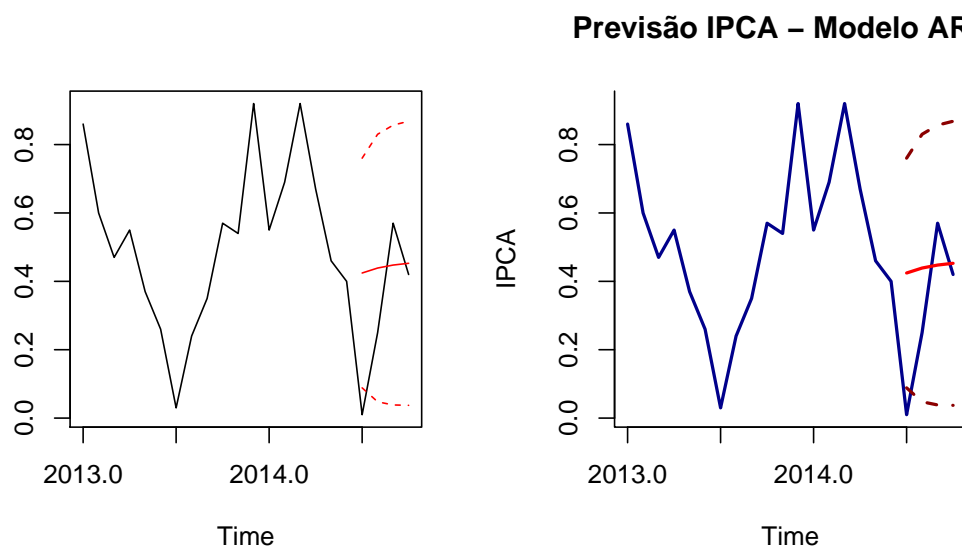
plot(window(ipca, start=2013),
      col='darkblue', lwd=2, bty='l', ylab='IPCA',
      main='Previsão IPCA - Modelo AR(1)')

lines(ts(prev2$pred, start=c(2014,07), freq=12),
      col='red', lwd=2)

lines(ts(prev2$pred+1.96*prev2$se,
        start=c(2014,07), freq=12),
      col='darkred', lwd=2, lty=2)

lines(ts(prev2$pred-1.96*prev2$se,
        start=c(2014,07), freq=12),
      col='darkred', lwd=2, lty=2)

```



### 10.3 Avaliando previsões

Para qualquer dado modelo temos que a nossa variável de interesse pode ser descrita como

$$Y_t = \hat{Y}_t + \hat{\varepsilon}_t$$

onde  $\hat{Y}_t$  é a estimativa de  $Y_t$  e  $\hat{\varepsilon}_t$  é um erro. Logo, o erro de previsão é dado por

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t.$$

As duas medidas mais utilizadas para avaliar a previsão de um modelo, ou o quanto estamos errando, são a raiz quadrada do erro quadrático médio (REQM) e o erro absoluto médio (EAM). A primeira medida, REQM, é dada pela fórmula abaixo

$$REQM = \sqrt{\frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{T}}$$

enquanto que a segunda medida, EAM, é dada por

$$EAM = \frac{\sum_{t=1}^T |\hat{\varepsilon}_t|}{T}.$$

Note primeiro que as duas medidas estão na mesma unidade que a variável de interesse,  $Y_t$ , logo podemos avaliar diretamente o impacto do erro. Em segundo lugar, veja a sutil diferença que existe entre essas duas medidas em relação ao que se busca medir efetivamente. Preferimos a primeira medida se queremos, em média, errar menos (ou ter uma previsão mais centrada no valor observado da variável). Por outro lado, preferimos a segunda medida quando buscamos um modelo que tenha, em média, valores mais baixos de erro (ou seja, quando erramos não queremos estar muito longe do valor observado)<sup>35</sup>. Obviamente vale ressaltar que a acurácia do modelo (segundo estas medidas) pode mudar dependendo do horizonte de previsão (denominado de  $T$  neste caso).

Vimos anteriormente que segundo o *AIC* os dois "melhores" modelos para o comportamento do IPCA são o *AR(1)* e o *ARMA(1,1)*. Podemos agora comparar estes dois modelos utilizando estas duas medidas de erro de previsão. Para isso precisamos construir uma série de erros de previsão (de um passo à frente). O processo que iremos seguir é dado a seguir.

1. Estimar modelos utilizando dados do IPCA referente ao período de janeiro de 2007 até dezembro de 2012;
2. Fazer previsão um passo à frente e salvar previsão;

---

<sup>35</sup>No caso do EAM todos os erros recebem o mesmo peso, enquanto que no caso do REQM erros menores receberem pesos menores.

3. Calcular e salvar erro de previsão;
4. Adicionar uma observação e reestimar modelos;
5. Repetir passos 2-4 até chegar ao final da amostra.

Seguindo a lógica descrita acima partimos para a parte prática. Primeiro construímos a série de erros de previsão.

```
# Modelo 1

erro1 <- matrix(NA, nrow=length(window(ipcad, start=2013)), ncol=1)

for (i in 1:length(erro1)){

  erro1[i] <- ipcad[(length(window(ipcad, end=c(2012, 12)))+i)] -

    predict(
      arima(ipcad[1:(length(window(ipcad, end=c(2012, 12)))+i-1]), order=c(1,0,0)),
      n.ahead=1)$pred

  erro1 <- ts(erro1, start=2013, freq=12)

}

# Modelo 2

erro2 <- matrix(NA, nrow=length(window(ipcad, start=2013)), ncol=1)

for (i in 1:length(erro1)){

  erro2[i] <- ipcad[(length(window(ipcad, end=c(2012, 12)))+i)] -

    predict(
```

```

    arima(ipcad[1:(length(window(ipcad, end=c(2012, 12)))+i-1)], order=c(1,0,1)),
    n.ahead=1)$pred

erro2 <- ts(erro2, start=2013, freq=12)

}

```

Em seguida, calculamos as duas medidas.

```

reqm1 <- sqrt(sum(erro1^2)/length(erro1))
reqm2 <- sqrt(sum(erro2^2)/length(erro2))

reqm1

## [1] 0.1973854

reqm2

## [1] 0.1985689

eam1 <- sum(abs(erro1))/length(erro1)
eam2 <- sum(abs(erro2))/length(erro2)

eam1

## [1] 0.1613305

eam2

## [1] 0.1591327

```

Assim temos que o REQM dos modelos 1 e 2 são 0.197 e 0.199, respectivamente. Já o EAM é de 0.161 e 0.159, respectivamente. Logo, temos que pela primeira medida o Modelo 1 se sai melhor, enquanto que pela segunda medida o Modelo 2 deveria ser o

escolhido<sup>36</sup>.

Como se trata de medidas usadas comumente, talvez seria interessante criar uma função que calcula o REQM e o EAM. Abaixo segue um exemplo de como se poderia criar essas funções.

```
# Função para calcular REQM

reqm <- function(serie){

  reqm <- sqrt(sum(serie^2)/length(serie))

  print(reqm)

}
```

```
# Função para calcular EAM

eam <- function(serie){

  eam <- sum(abs(serie))/length(serie)

  print(eam)

}
```

---

<sup>36</sup>É claro que estas duas medidas não são as únicas maneiras de se comparar o poder de previsão de dois modelos, porém para o intuito deste curso somente abordaremos esta duas medidas. Os mais interessados podem procurar testes como o de *Granger-Newbold*, *Diebold-Mariano* e a estatística *U* de *Thiel*.

## 11 Modelos multivariados

Nas seções anteriores estudamos modelos onde as variáveis de interesse eram explicadas pelos seus valores passados. Porém, talvez seja de nosso interesse incluir outras variáveis explicativas no modelo. No caso da inflação medida pelo IPCA, por exemplo, talvez seria interessante incluir a taxa de desemprego dado que acreditamos na Lei de Okun. Vejamos agora o caso de modelos multivariados, onde usamos outras variáveis além de defasagens da variável endógena.

### 11.1 Modelos dinâmicos com variáveis estacionárias

Aqui surge mais uma diferença entre a econometria de *cross-sections* e a de séries temporais, os efeitos que uma variável tem sobre a outra é dinâmico e existe diferença entre efeitos de curto e longo prazo. De forma a ilustrar assumo o processo auto-regressivo com defasagens distribuídas,  $ADL^{37}$ , de duas variáveis a seguir

$$Y_t = \mu + \theta_1 Y_{t-1} + \phi_0 X_t + \phi_1 X_{t-1} + \varepsilon_t.$$

Este modelo também é chamado de um modelo  $ADL(1)$ , dado que a maior ordem das defasagens é 1. Consequentemente, um modelo  $ADL(p)$  pode ser escrito como

$$Y_t = \mu + \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p} + \phi_0 X_t + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t.$$

Obviamente, modelos ADL podem ser estendidos de forma a incluir mais regressores e componentes determinísticos (como por exemplo variáveis *dummies*). Como na econometria básica, o multiplicador de impacto das variáveis são dadas pelas derivadas parciais, ou seja, o multiplicador de impacto, ou efeito "instantâneo", do modelo  $ADL(1)$  acima é dado por

$$\frac{\partial Y_t}{\partial X_t} = \phi_0.$$

Os efeitos de longo prazo, por outro lado, podem ser obtidos de maneira recursiva como a seguir

$$Y_{t+1} = \mu + \theta_1 Y_t + \phi_0 X_{t+1} + \phi_1 X_t + \varepsilon_{t+1}$$

---

<sup>37</sup>Como comumente denotada, do inglês "Autoregressive Distributed Lag".

$$\frac{\partial Y_{t+1}}{\partial X_t} = \theta_1 \frac{\partial Y_t}{\partial X_t} + \phi_1 = \theta_1 \phi_0 + \phi_1$$

$$Y_{t+2} = \mu + \theta_1 Y_{t+1} + \phi_0 X_{t+2} + \phi_1 X_{t+1} + \varepsilon_{t+2}$$

$$\frac{\partial Y_{t+2}}{\partial X_t} = \theta_1 \frac{\partial Y_{t+1}}{\partial X_t} + \phi_1 = \theta_1(\phi_0 + \phi_1)$$

...

Logo, somando todos esses efeitos teremos que o efeito de longo prazo é dado por

$$\begin{aligned} \sum_{i=0}^{\infty} \theta_1^i \phi_0 + \sum_{i=0}^{\infty} \theta_1^i \phi_1 + \theta_1^{\infty} \phi_1 \\ = \frac{\phi_0 + \phi_1}{1 - \theta_1} \end{aligned}$$

Uma maneira alternativa de se obter o multiplicador de longo prazo é através da imposição da condição de equilíbrio de longo prazo, isto é, que as variáveis se mantenham no mesmo nível no equilíbrio. Ou seja, que  $E(Y_t) = E(Y_{t-1})$  e  $E(X_t) = E(X_{t-1})$ . Desta maneira temos que

$$E(Y_t) = E(\mu + \theta_1 Y_{t-1} + \phi_0 X_t + \phi_1 X_{t-1} + \varepsilon_t).$$

$$E(Y_t) - \theta_1 E(Y_{t-1}) = \mu + \phi_0 E(X_t) + \phi_1 E(X_{t-1})$$

$$(1 - \theta_1)E(Y_t) = \mu + (\phi_0 + \phi_1)E(X_t)$$

$$E(Y_t) = \frac{\mu}{(1 - \theta_1)} + \frac{\phi_0 + \phi_1}{1 - \theta_1} E(X_t)^{38}$$

Finalmente, temos que no longo prazo

$$\frac{\partial Y_t}{\partial X_t} = \frac{\phi_0 + \phi_1}{1 - \theta_1}.$$

---

<sup>38</sup>Note que podemos denotar esta equação como  $Y_t = \alpha + \beta X_t$ , onde  $\beta$  terá uma interpretação de longo prazo.

## 11.2 Modelos de correção de erro

Podemos agora reescrever o modelo ADL(1) anterior como um modelo de correção de erro. Para isso, devemos subtrair  $Y_{t-1}$  e  $\phi_0 X_{t-1}$  de ambos os lados, obtendo

$$\Delta Y_t = \mu - (1 - \theta_1)Y_{t-1} + \phi_0 \Delta X_t + (\phi_0 + \phi_1)X_{t-1} + \varepsilon_t$$

ou

$$\Delta Y_t = \phi_0 \Delta X_t - (1 - \theta_1)[Y_{t-1} - \alpha - \beta X_{t-1}] + \varepsilon_t$$

onde  $\alpha \equiv \mu/(1 - \theta_1)$ ,  $\beta \equiv (\phi_0 + \phi_1)/(1 - \theta_1)$ ,  $(1 - \theta_1)$  é o parâmetro de ajuste e  $Y_{t-1} - \alpha - \beta X_{t-1}$  é o erro de equilíbrio.

## 11.3 Estimando e testando

Modelos ADL simples podem ser estimados de maneira consistente utilizando MQO e aqui utilizaremos a estratégia de modelar do caso mais geral para um mais específico. Esta será a estratégia utilizada principalmente por um motivo: a omissão de variáveis causa problemas (viés) maiores que a inclusão de variáveis não significantes (ineficiência/perda de graus de liberdade). Os passos a seguir são, em geral,

1. Formular e estimar um modelo *ADL* irrestrito.
2. Testar sucessivamente significância de regressores.
3. Testes diagnósticos dos resíduos do modelo final.

Onde como sempre, os testes de diagnósticos devem incluir testes para autocorrelação (*Box-Pierce*, *Ljung-Box* e *Breusch-Godfrey*, por exemplo), heterocedasticidade (*White*, *Breusch-Pagan / LM* e *Ljung-Box* nos resíduos ao quadrado, por exemplo) e normalidade (*Jarque-Bera* e *Shapiro-Wilk*, por exemplo).

## 11.4 Causalidade de Granger

Se valores passados de  $X$  nos ajudam a prever  $Y$ , então diz-se que  $X$  *Granger* causa  $Y$ , ou simplesmente  $X \rightarrow Y$ . Assuma a seguinte equação que descreve a relação entre  $X$  e  $Y$ :

$$Y_t = \sum_1^n \alpha_i X_{t-i} + \sum_1^n \beta_i Y_{t-i} + \varepsilon_{1t}.$$



Se  $\sum_1^n \alpha_i \neq 0$ , ou seja, se todos  $\alpha_i$  são conjuntamente diferentes de zero, então temos que  $X \rightarrow Y$ . Porém, o contrário também pode valer, isto é  $Y \rightarrow X$ . Trocando a causalidade temos então a seguinte relação

$$X_t = \sum_1^n \gamma_i X_{t-i} + \sum_1^n \rho_i Y_{t-i} + \varepsilon_{2t}.$$

Onde, se valer que  $\sum_1^n \rho_i \neq 0$ , poderemos concluir que  $Y \rightarrow X$ .

## 11.5 Propriedades dos estimadores de MQO: Amostra Finita

Antes de continuar com a análise de modelos multivariados, vejamos algumas das hipóteses que devem ser satisfeitas de modo a garantir o "bom" funcionamento dos estimadores MQO.

### 11.5.1 Estimadores não viesados

#### H1 Linear nos parâmetros

#### H2 Média condicional zero

$$E(\varepsilon_t|X) = 0, \forall t = 1, \dots, n.$$

Note que para **H2** valer,  $\varepsilon_t$  deve ser não correlacionado com  $X_s$  mesmo quando  $s \neq t$ . Isto somente valerá quando todos os regressores forem estritamente exógenos. Esta hipótese pode ser bem difícil de garantir e problemas como omissão de variáveis e defasagens, erro de medida, causalidade reversa e até mesmo o caso de um modelo  $AR(1)$  violam esta hipótese.

#### H3 Ausência de colinearidade perfeita

Se estas hipóteses acima são válidas então temos que os estimadores MQO serão não viesados e, conseqüentemente,  $E(\hat{\beta}) = \beta$ .

### 11.5.2 Estimadores eficientes

**H4 Homocedasticidade** A variância de  $\varepsilon_t$  dado  $X$  deve ser o mesmo para todo  $t$ , ou seja

$$Var(\varepsilon_t|X) = Var(\varepsilon_t) = \sigma^2, t = 1, \dots, n.$$

**H5 Ausência de autocorrelação** A correlação dos erros entre dois períodos distintos dado  $X$  é zero,

$$\text{Corr}(\varepsilon_t, \varepsilon_s | X) = 0 \quad \forall t \neq s.$$

**Teorema de Gauss-Markov** Sob as hipóteses **H1-H5** os estimadores MQO são os melhores estimadores lineares não viesados.

### 11.5.3 Inferência

**H6 Normalidade** Os erros  $\varepsilon_t$  são independentes de  $X$  e  $\stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ .

Sob as hipóteses **H1-H6** os estimadores de MQO são normalmente distribuídos dado  $X$ . Além disso, sob a hipótese nula, as estatísticas  $t$  e  $F$  seguem as distribuições  $t$  e  $F$ , respectivamente, e a construção de intervalos de confiança são válidos.

## 11.6 Propriedades assintóticas dos estimadores de MQO

### H1 Dependência linear e fraca

#### Média condicional zero

$$E(\varepsilon_t | x_t) = 0, \quad \forall t = 1, \dots, n.$$

Note que aqui a exogeneidade exigida é somente contemporânea, ou seja, uma hipótese consideravelmente mais fraca.

#### Ausência de colinearidade perfeita

Sob as hipóteses **H1-H3** os estimadores de MQO são consistentes, ou seja,  $\hat{\beta} \rightarrow \beta$ .

**H4 Homocedasticidade** Os erros são contemporaneamente homocedásticos.

$$\text{Var}(\varepsilon_t | x_t) = \sigma^2.$$

#### H5 Ausência de autocorrelação

$$\text{Corr}(\varepsilon_t, \varepsilon_s | x_t, x_s) = 0.$$

**Normalidade assintótica dos estimadores de MQO** Sob as hipóteses **H1-H5** os estimadores de MQO são assintoticamente normalmente distribuídos. Consequentemente, os erros, estatísticas  $t$ ,  $F$  e  $LM$  são assintoticamente válidos.

## 11.7 Modelos dinâmicos com variáveis não estacionárias

Não estacionaridade causa estimadores de MQO inconsistentes e, conseqüentemente, os resultados das regressões não podem ser confiados. Vejamos a seguir o problema do uso de variáveis não estacionárias.

### 11.7.1 Regressão espúria

Suponha duas variáveis aleatórias  $X_t$  e  $Y_t$  caracterizadas por um passeio aleatório. Podemos representá-las como

$$X_t = X_{t-1} + \varepsilon_{Xt}$$

$$Y_t = Y_{t-1} + \varepsilon_{Yt}$$

onde  $\varepsilon_{Xt}, \varepsilon_{Yt} \stackrel{i.i.d.}{\sim} (0, 1)$ . Se rodarmos a regressão a seguir

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

iremos, em geral, ter como resultado um  $R^2$  relativamente alto e um  $\beta$  estatisticamente significativo<sup>39</sup>. Ou seja, iremos concluir, erroneamente, que existe uma relação entre essas duas variáveis.

De modo a exemplificar iremos agora simular duas variáveis não estacionárias e fazer a regressão de uma na outra.

```
set.seed(3)

x <- cumsum(rnorm(200))
y <- cumsum(rnorm(200))
```

Como conhecemos o processo que gera as duas séries, temos certeza absoluta de que não há nenhuma relação estatística entre elas. Rodando a regressão obtemos:

```
summary(lm(y~x))
```

---

<sup>39</sup>Uma estatística *Durbin-Watson* baixa também provavelmente será encontrada.

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2269  -2.8512  -0.1129   2.9624  12.4666
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.3356     0.3613  23.073  < 2e-16 ***
## x             0.7510     0.0959   7.831 2.88e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.061 on 198 degrees of freedom
## Multiple R-squared:  0.2365, Adjusted R-squared:  0.2326
## F-statistic: 61.32 on 1 and 198 DF,  p-value: 2.883e-13
```

Apesar do  $R^2$  não ser tão alto, especialmente em se tratando de séries temporais, pode-se notar que o coeficiente de  $X_t$  é altamente significativo.

Agora vamos ver o que acontece quando acrescentamos um *drift* nas séries<sup>40</sup>.

```
driftx <- 0.5
drifty <- 0.2
xd <- cumsum(driftx + rnorm(200))
yd <- cumsum(drifty + rnorm(200))

summary(lm(yd~xd))
```

---

<sup>40</sup>Note que utilizaremos a mesma "semente" (ou o mesmo número para a opção `set.seed`) e, portanto, estaremos analisando as mesmas séries porém com *drift*.

```
##
## Call:
## lm(formula = yd ~ xd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.2085 -1.3834  0.2746  1.4287  7.2029
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.650790   0.431231  -13.10  <2e-16 ***
## xd           0.207911   0.006627   31.37  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.591 on 198 degrees of freedom
## Multiple R-squared:  0.8325, Adjusted R-squared:  0.8317
## F-statistic: 984.2 on 1 and 198 DF,  p-value: < 2.2e-16
```

Como podemos ver, a existência do *drift* no processo que gera os dados agrava o problema da regressão espúria.

## 11.8 Cointegração

Vamos agora assumir que exista de fato uma relação entre  $X_t$  e  $Y_t$ , como a seguir<sup>41</sup>

$$X_t = X_{t-1} + \varepsilon_{Xt}$$

$$Y_t = \beta X_t + \varepsilon_{Yt}.$$

A relação entre  $X_t$  e  $Y_t$  é linear e pode ser caracterizada pelo vetor  $[1, -\beta]$ . Este vetor é comumente chamado de vetor cointegrante, isto porque ele é o vetor que gera uma combinação estacionária de  $X_t$  e  $Y_t$ .

---

<sup>41</sup> Assuma as mesmas propriedades anteriores de  $\varepsilon_{Xt}$  e  $\varepsilon_{Yt}$ .

**Definição** As variáveis  $X_t$  e  $Y_t$  são ditas cointegradas se existe pelo menos uma combinação linear entre elas,  $\varepsilon_{Yt} = Y_t - \beta X_t$ , que gere uma variável estacionária, ou seja,  $I(0)$ .

Este conceito de cointegração tem uma ligação estreita com o conceito de equilíbrio de longo prazo. Vejamos agora a seguinte relação:

$$Y_t = \alpha + \beta X_t + u_t$$

e defina o erro de equilíbrio como  $u_t = Y_t - \alpha - \beta X_t$ .  $u_t$  irá nos dizer o quão longe  $Y_t$  está de seu valor de equilíbrio, dado  $X_t$ . Se  $u_t$  for estacionário e flutuando ao redor de zero, então o sistema estará, em média, em equilíbrio, apesar de  $Y_t$  e  $X_t$  serem não estacionários. Caso contrário, se  $u_t$  for não estacionário o erro de equilíbrio poderá ficar cada vez maior.

Agora que vimos as definições básicas de regressão espúria e cointegração podemos prosseguir para a parte formal de como distinguir entre uma relação espúria entre duas variáveis e uma relação de cointegração. Como podemos imaginar, a grande diferença está nos resíduos resultantes da regressão entre as duas variáveis de interesse. Logo, para descobrir se temos uma relação de cointegração devemos testar se os resíduos são estacionários. Assim, assumindo que  $Y_t$  e  $X_t$  são  $I(1)$ :

1. Rode a regressão

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

2. Teste se  $\hat{\varepsilon}_t$  é  $I(0)$  ou  $I(1)$ .

Se  $\hat{\varepsilon}_t$  for  $I(1)$  então  $X_t$  e  $Y_t$  não são cointegrados. Porém, se  $\hat{\varepsilon}_t$  for  $I(0)$  não podemos rejeitar a hipótese de que  $X_t$  e  $Y_t$  são cointegrados<sup>42</sup>.

## 11.9 Cointegração e modelos de correção de erro

**Teorema de representação de Granger** Se um conjunto de variáveis são cointegradas, então existe uma representação de correção de erros válida dos dados

$$\Delta Y_t = \gamma_0 + \gamma_1 \Delta X_t + \gamma_2 u_{t-1} + \varepsilon_t$$

---

<sup>42</sup>Note que a maneira mais comum de se testar esta hipótese é através dos testes *ADF*, porém neste caso não são utilizados os mesmo valores críticos.

onde  $\varepsilon_t$  é um ruído branco e

$$u_{t-1} = Y_{t-1} - \alpha - \beta X_{t-1}$$

é o termo de erro de equilíbrio derivado da equação de cointegração, ou seja, da equação de equilíbrio de longo prazo

$$Y_t = \alpha + \beta X_t + u_t.$$

Se  $X_t$  e  $Y_t$  são  $I(1)$  e possuem uma relação de longo prazo, então algo deve trazer os erros de equilíbrio de volta para zero. Este algo é justamente o erro de equilíbrio do modelo de correção de erro.

### 11.10 Estimando um modelo de correção de erro usando o método de dois passos de *Engle-Granger*

Tome como exemplo a função de consumo keynesiana

$$C = f(Y) = \alpha + \beta Y.$$

Assumimos comumente que quando a renda aumenta o consumo aumenta, porém em proporção menor, ou seja

$$0 < \frac{\partial C}{\partial Y} = \beta < 1.$$

Assumimos frequentemente também que a propensão marginal a consumir (PMC) de curto prazo é menor que a PMC de longo prazo,  $\beta$  neste caso, devido a formação de hábito. Vejamos agora um exemplo prático utilizando as séries trimestrais de consumo e renda disponível retiradas do Ipeadata. Ambas as séries começam no primeiro trimestre de 1991, porém utilizaremos somente os dados a partir de 1996. Abaixo importamos e visualizamos as séries.

```
coint <- read.csv("coint.csv",
                  sep=";", dec=",", stringsAsFactors=FALSE)
coint <- ts(coint, start=1991, freq=4)
coint <- window(coint[,-1], start=1996)

cons <- coint[,1]
```

```

renda <- coint[,2]

plot(cons, col='darkblue', bty='l',
     xlab='', ylab='R$', ylim=range(coint))

par(new=T)

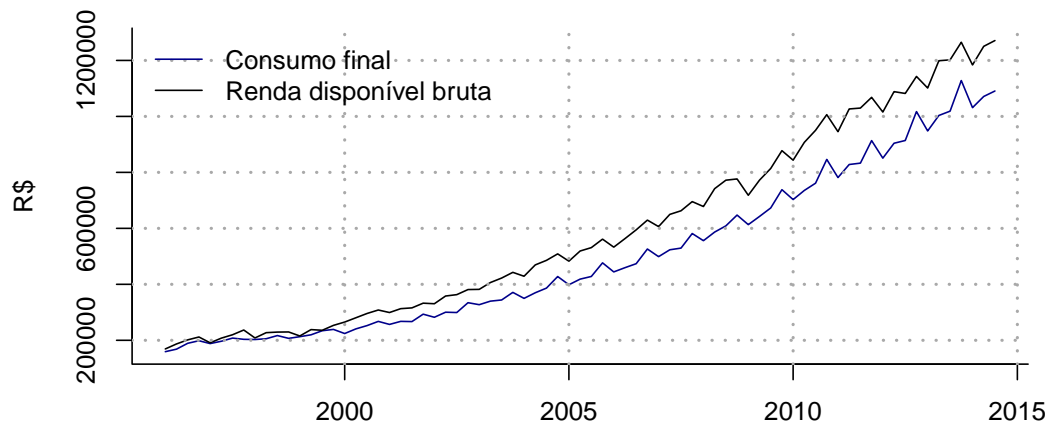
plot(renda, col='black',
     ann=F, axes=F, ylim=range(coint))

par(new=F)

grid(col='darkgrey', lwd=2)

legend('topleft', c('Consumo final', 'Renda disponível bruta'),
      col=c('darkblue', 1), bty='n', lty=1)

```



Assumimos anteriormente que  $Y_t$  e  $X_t$  eram  $I(1)$ , isto porque não faria sentido buscar cointegração entre variáveis que são estacionárias. Apesar das séries apresentarem uma tendência clara, faremos um teste formal para verificar que ambas as séries são integradas



de mesma ordem. Para isto utilizaremos o teste  $ADF$ <sup>43</sup>.

```
ur.df(cons)

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: 6.1682

ur.df(renda)

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: 6.7522

ur.df(diff(cons))

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -5.3537

ur.df(diff(renda))

##
## #####
```

---

<sup>43</sup>Outra maneira seria utilizar o teste de cointegração de *Durbin-Watson*, por exemplo.

```
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -4.34
```

Dados os resultados acima podemos concluir que as séries de consumo e renda são  $I(1)$ . Logo, prosseguimos com a regressão que estima a função consumo.

```
reg <- lm(cons~renda)
summary(reg)

##
## Call:
## lm(formula = cons ~ renda)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40789 -12108   -622   11385   67986
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.823e+03  4.818e+03   1.416   0.161
## renda        8.326e-01  7.046e-03 118.170 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20590 on 73 degrees of freedom
## Multiple R-squared:  0.9948, Adjusted R-squared:  0.9947
## F-statistic: 1.396e+04 on 1 and 73 DF,  p-value: < 2.2e-16
```

Os resultados preliminares indicam que nossa hipótese sobre a PMC de longo prazo é válida, ou seja, está entre 0 e 1. Agora devemos descobrir se esta regressão é espúria ou se realmente ela nos dá algo de valor. Para isso devemos salvar os resíduos e verificar se são estacionários ou não.

```
ur.df(reg$resid)

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -3.7513
```

Como foi dito anteriormente, apesar de utilizarmos o teste *ADF* não podemos utilizar os valores críticos normalmente utilizados. Para o teste de cointegração utilizaremos os valores críticos disponibilizados na Tabela 2 do arquivo [dfdistrib7b.pdf](#). Como podemos ver, rejeitamos a hipótese de raiz unitária, logo os resíduos da regressão são estacionários e podemos concluir a existência de cointegração entre as variáveis de interesse. Logo, o coeficiente  $\beta$  de valor 0.83, pode ser interpretado como a PMC de longo prazo. Agora, podemos prosseguir e estimar o modelo de correção de erro utilizando os resíduos do primeiro passo. Ou seja, iremos estimar a seguinte equação

$$\Delta Cons_t = \gamma_0 + \gamma_1 \Delta Renda_t + \gamma_2 \hat{u}_{t-1} \varepsilon_t$$

onde  $\hat{u}_t$  são os resíduos do primeiro passo.

```
u <- reg$resid[-length(reg$resid)]
dCons <- diff(cons)
dRenda <- diff(renda)

ecm <- lm(dCons~dRenda + u)
```

```
summary(ecm)

##
## Call:
## lm(formula = dCons ~ dRenda + u)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -28684 -11416  -2883   11382   60992
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -23.03217  2134.84957  -0.011  0.991422
## dRenda        0.83767    0.06075   13.788  < 2e-16 ***
## u            -0.39301    0.10030   -3.919  0.000203 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16670 on 71 degrees of freedom
## Multiple R-squared:  0.7855, Adjusted R-squared:  0.7795
## F-statistic: 130 on 2 and 71 DF,  p-value: < 2.2e-16
```

Primeiramente, devemos notar que o coeficiente do termo de correção de erro,  $\gamma_1$ , foi estimado ter o valor de  $-0.39$ , ou seja, ele é negativo. Isto deve sempre ser verdade, pois significa que quando existe um erro positivo o modelo ajusta-se para "baixo". Ou seja, o termo de correção de erro está fazendo exatamente o que ele deveria fazer, ajustar o modelo de volta para o equilíbrio. Este valor nos dá justamente a "velocidade" de convergência ou velocidade de ajuste do modelo.

Além disso, podemos ver que a PMC de curto prazo foi estimada em 0.84, indo contra nossa hipótese de formação de hábito.

### 11.11 Estimando um modelo de correção de erro em um passo

Acabmos de ver um método de se estimar um modelo de correção de erros. Infelizmente, ao se usar dois passos este tipo de estimação torna-se menos eficiente, isto é, seus estimadores possuem um desvio-padrão maior. Uma maneira de contornar este problema é de se estimar o modelo de correção diretamente. Em outras palavras, queremos estimar a seguinte equação

$$\Delta Cons_t = \alpha + \gamma_1 \Delta Renda_t + \phi_1 Cons_{t-1} + \phi_2 Renda_{t-1} + \epsilon_t.$$

```
cCons <- cons[-length(cons)]
cRenda <- renda[-length(renda)]

ecm2 <- lm(dCons~dRenda + cCons + cRenda)
```

```
summary(ecm2)

##
## Call:
## lm(formula = dCons ~ dRenda + cCons + cRenda)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28785 -12032  -2679   10180   59321
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  960.69728 4097.08048   0.234 0.815294
## dRenda        0.83410    0.06149  13.565 < 2e-16 ***
## cCons       -0.39302    0.10083  -3.898 0.000220 ***
## cRenda        0.33023    0.08416   3.924 0.000201 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16760 on 70 degrees of freedom
## Multiple R-squared:  0.7863, Adjusted R-squared:  0.7771
## F-statistic: 85.86 on 3 and 70 DF,  p-value: < 2.2e-16
```

Como podemos ver, obtemos aproximadamente o mesmo valor para a PMC de curto prazo, ou seja, 0.83. Além disso temos que os resíduos dessa regressão são estacionários:

```
ur.df(ecm2$resid)

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -7.8055
```

A PMC de longo prazo pode ser obtido pelo negativo da razão entre os coeficientes das variáveis em nível, neste caso

$$\text{PMC de longo prazo} = -\frac{0.33}{-0.39} = 0.84$$

## 11.12 Um pouco mais sobre cointegração

No exemplo anterior deveríamos ser capazes de reverter a ordem das variáveis, ou seja, colocar renda como variável explicada, e obter os mesmos resultados. Vejamos os resultados desta regressão.

```
reg2 <- lm(renda~cons)
```

```
summary(reg2)

##
## Call:
## lm(formula = renda ~ cons)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -77741 -14730    761  15201  49655
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.058e+03  5.820e+03  -0.869    0.388
```

```
## cons          1.195e+00  1.011e-02 118.170    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24660 on 73 degrees of freedom
## Multiple R-squared:  0.9948, Adjusted R-squared:  0.9947
## F-statistic: 1.396e+04 on 1 and 73 DF,  p-value: < 2.2e-16
```

Notem que obtemos os mesmos valores para as estatísticas  $t$  e  $R^2$ . O inverso do coeficiente obtido neste caso, também se aproxima bastante do valor obtido na primeira regressão que rodamos, onde no primeiro caso obtemos 0.833 e o inverso do coeficiente desta regressão é 0.837.

### 11.12.1 Cointegração e Causalidade de *Granger*

Por definição, se duas variáveis  $I(1)$  são cointegradas, então pelo menos uma das variáveis deve *Granger* causar a outra. O problema de cointegração se torna mais complexa quando o número de variáveis na relação de longo prazo aumenta (especialmente se as ordens de integração são diferentes). Assuma, por exemplo, que  $y_t$  seja  $I(0)$  enquanto que  $x_t$  e  $z_t$  sejam  $I(1)$  e que exista a seguinte relação

$$y_t = \alpha + \beta x_t + \gamma z_t + \varepsilon_t.$$

Dada esta relação temos que

$$\varepsilon_t \sim I(0) \text{ se, somente se } (\beta x_t + \gamma z_t) \sim I(0).$$

Se as variáveis são integradas de ordens diferentes, então devem existir na equação no mínimo duas variáveis de cada ordem de integração. Uma variável  $I(1)$ , por exemplo, nunca poderá ser cointegrada com uma variável  $I(2)$ .

O problema da metodologia de *Engle-Granger* é que ela nos dá no máximo um vetor de cointegração, independente do número de variáveis envolvidas. Porém, se  $n$  variáveis estão envolvidas no processo, então podem existir até  $n - 1$  vetores cointegrantes. O vetor cointegrante estimado no caso de 3 ou mais variáveis utilizando esta metodologia pode não ser único. Podendo ser uma combinação de diferentes vetores cointegrantes.

A metodologia de *Johansen* nos diz quantos vetores cointegrantes existem entre  $n$  variáveis. Voltaremos a este mais à frente após introduzir o tópico de modelos de vetores autoregressivos.



## Referências

- Dickey, D. A. and Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.
- Enders, W. *Applied Econometric Times Series*. Wiley Series in Probability and Statistics. Wiley, 2009.
- Pfaff, B. *Analysis of integrated and cointegrated time series with R*. Springer, New York, second edition, 2008.