

# Análise de Séries Temporais

Fernando Lucambio

Departamento de Estatística  
Universidade Federal do Paraná

Dezembro, 2020

Nas Parte I e Parte II vimos que se  $X_t$  é um passeio aleatório,  $X_t = X_{t-1} + W_t$ , então por diferenciação de  $X_t$ , achamos que  $\nabla X_t = W_t$  é estacionário. Em muitas situações, séries temporais podem ser consideradas como sendo compostas por dois componentes, um componente de tendência não estacionário e um componente estacionário de média zero. Por exemplo, na Seção II.1, consideramos o modelo

$$X_t = \mu_t + Y_t,$$

onde  $\mu_t = \beta_0 + \beta_1 t$  e  $Y_t$  é estacionário. Diferenciar esse processo levará a um processo estacionário:

$$\nabla X_t = X_t - X_{t-1} = \beta_1 + Y_t - Y_{t-1} = \beta_1 + \nabla Y_t.$$

O modelo ARMA integrado ou ARIMA é uma ampliação da classe de modelos ARMA para incluir a diferenciação.

### Definição III.11. Modelo ARIMA.

O processo  $X_t$  é dito ser  $ARIMA(p, d, q)$  se

$$\nabla^d X_t = (1 - B)^d X_t,$$

é  $ARMA(p, q)$ . Em geral, vamos escrever o modelo como

$$\phi(B)(1 - B)^d X_t = \theta(B)W_t.$$

Se  $E(\nabla^d X_t) = \mu$ , escrevemos o modelo como

$$\phi(B)(1 - B)^d X_t = \delta + \theta(B)W_t,$$

onde  $\delta = \mu(1 - \phi_1 - \dots - \phi_p)$ .

Deve ficar claro que, como  $Y_t = \nabla^d X_t$  é ARMA, podemos usar os métodos da Seção III.4 para obter previsões de  $Y_t$ , que, por sua vez, levam a previsões para  $X_t$ . Por exemplo, se  $d = 1$ , dadas as previsões  $Y_{n+m}^n$  para  $m = 1, 2, \dots$ , temos  $Y_{n+m}^n = X_{n+m}^n - X_{n+m-1}^n$ , de modo que

$$X_{n+m}^n = Y_{n+m}^n + X_{n+m-1}^n,$$

com condição inicial  $X_{n+1}^n = Y_{n+1}^n + X_n$ , notando que  $X_n^n = X_n$ .

É um pouco mais difícil obter os erros de previsão  $P_{n+m}^n$ , mas para grandes  $n$ , a aproximação usada na Seção III.4, funciona bem. Ou seja, o erro de previsão do quadrado médio pode ser aproximado por

$$P_{n+m}^n = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^{*2},$$

sendo  $\psi_j^{*2}$  o coeficiente de  $z^j$  em  $\psi^*(z) = \theta(z)/(\phi(z)(1-z)^d)$ .

### Exemplo III.37. Passeio aleatório com tendência.

Para fixar ideias, começamos considerando o passeio aleatório com modelo de tendência apresentado primeiro no Exemplo I.11, ou seja,

$$X_t = \delta + X_{t-1} + W_t,$$

para  $t = 1, 2, \dots$  e  $X_0 = 0$ . Tecnicamente, o modelo não é *ARIMA*, mas podemos incluí-lo trivialmente como um modelo *ARIMA*(0, 1, 0). Dados  $X_1, \dots, X_n$ , a previsão de um passo à frente é dada por

$$X_{n+1}^n = E(X_{n+1} | X_n, \dots, X_1) = \delta + X_n.$$

A previsão de dois passos é dada por  $X_{n+2}^n = \delta + X_{n+1}^n = 2\delta + X_n$  e, conseqüentemente, a previsão de  $m$ -passos à frente, para  $m = 1, 2, \dots$  é

$$X_{n+m}^n = m\delta + X_n.$$

### Exemplo III.38. *IMA(1, 1)* e *EWMA*.

O modelo *ARIMA(0, 1, 1)* ou *IMA(1, 1)* é de interesse porque muitas séries temporais econômicas podem ser modeladas com sucesso dessa maneira. Além disso, o modelo leva a um método de previsão frequentemente usado e abusado, chamado de médias móveis exponencialmente ponderadas ou *EWMA* (Exponentially Weighted Moving Averages).

Vamos escrever o modelo como

$$X_t = X_{t-1} + W_t - \lambda W_{t-1},$$

com  $|\lambda| < 1$ , para  $t = 1, 2, \dots$  e  $X_0 = 0$ , porque este modelo é mais fácil de se trabalhar aqui e leva à representação padrão para *EWMA*.

## Método Holt-Winters

Este método de previsão é popular porque é fácil de usar, precisamos apenas reter o valor da previsão anterior e a observação atual para prever o próximo período de tempo.

Infelizmente, como sugerido anteriormente, o método é frequentemente abusado porque alguns analistas não verificam se as observações seguem um processo  $IMA(1,1)$  e muitas vezes arbitrariamente escolhem valores de  $\lambda$ .

A seguir, mostramos como gerar 100 observações de um modelo  $IMA(1,1)$  com  $\lambda = -\theta = 0.8$  e depois calcular e exibir o  $EWMA$  ajustado sobreposto aos dados. Isso é feito usando o comando **Holt-Winters** em R, veja o arquivo de ajuda **?HoltWinters** para detalhes:

```
> set.seed(666)
> x = arima.sim(list(order = c(0,1,1), ma = -0.8), n = 100)
> (x.ima = HoltWinters(x, beta=FALSE, gamma=FALSE))
Holt-Winters exponential smoothing without trend and
                                without seasonal component.
```

Call:

```
HoltWinters(x = x, beta = FALSE, gamma = FALSE)
```

Smoothing parameters:

```
alpha: 0.1663072
```

```
beta : FALSE
```

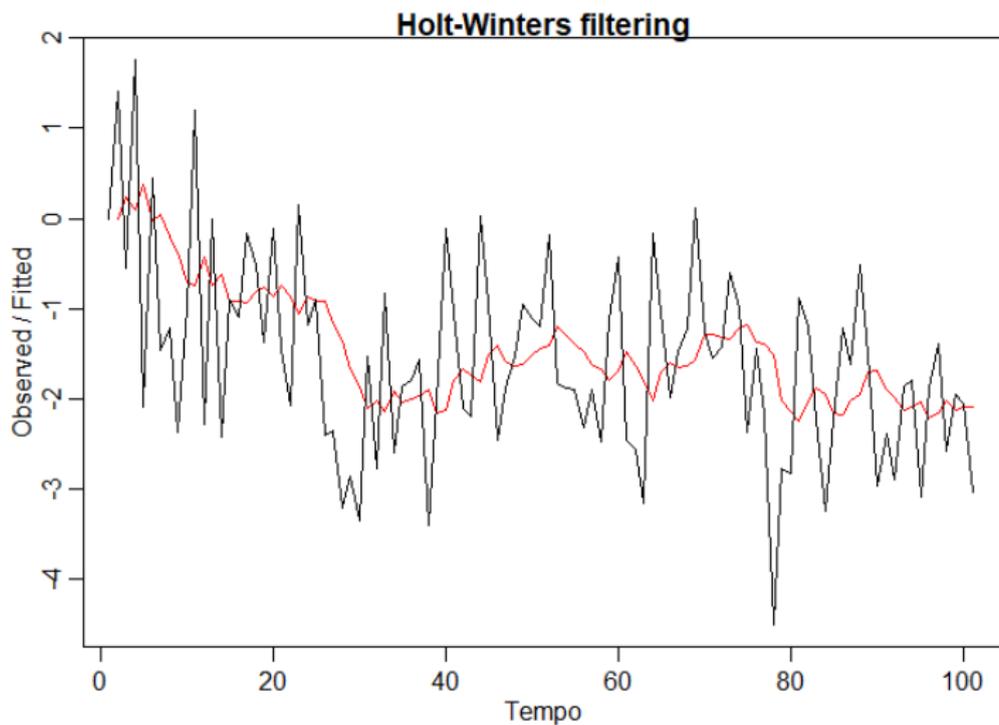
```
gamma: FALSE
```

Coefficients:

```
    [,1]
```

```
a -2.241533
```

```
> plot(x.ima, xlab="Tempo")
```



Existem algumas etapas básicas para ajustar modelos *ARIMA* a dados de séries temporais. Essas etapas envolvem:

- ▶ plotando os dados,
- ▶ possivelmente transformando os dados,
- ▶ identificar as ordens de dependência do modelo,
- ▶ estimação dos parâmetros,
- ▶ diagnósticos e
- ▶ escolha do modelo.

Primeiro, como em qualquer análise de dados, devemos construir um gráfico de tempo dos dados e inspecionar o gráfico em busca de quaisquer anomalias. Se, por exemplo, a variabilidade nos dados aumenta com o tempo, será necessário transformar os dados para estabilizar a variância. Nesses casos, a classe de transformações de potência Box-Cox, poderia ser empregada. Além disso, pode-se sugerir uma transformação apropriada.

Em exemplos os dados se comportam como  $X_t = (1 + p_t)X_{t-1}$ , onde  $p_t$  é uma pequena alteração percentual do período  $t - 1$  para  $t$ , que pode ser negativo. Se  $p_t$  é um processo relativamente estável, então  $\nabla \log(X_t) \approx p_t$  será relativamente estável. Frequentemente,  $\nabla \log(X_t)$  é chamado retorno ou taxa de crescimento.

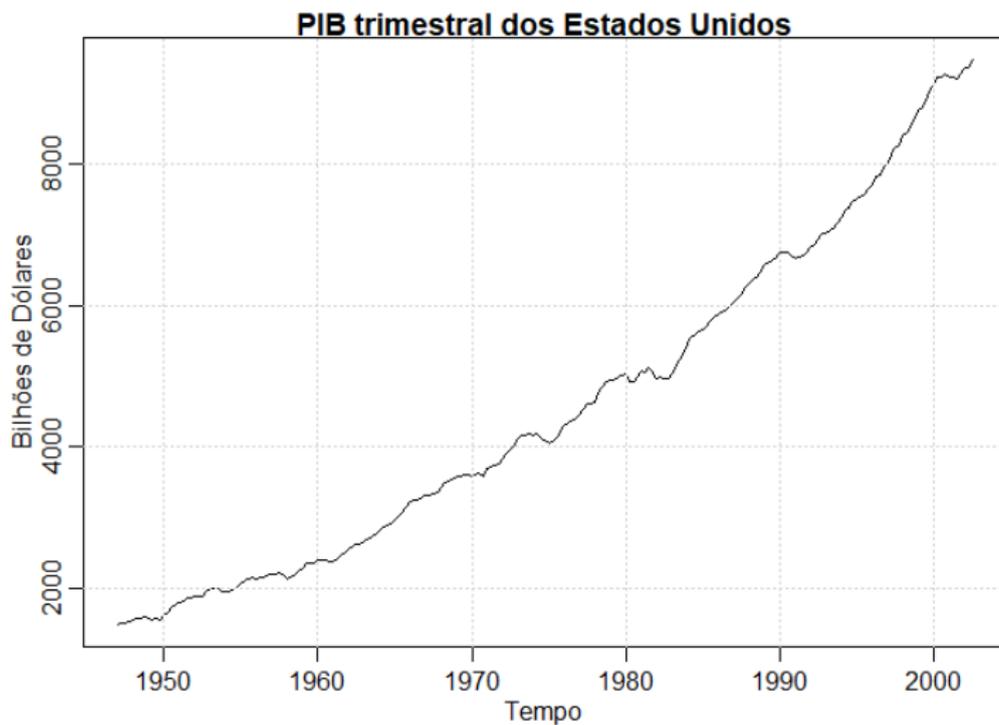
Depois de transformar adequadamente os dados, o próximo passo é identificar os valores preliminares da ordem autoregressiva  $p$ , a ordem de diferenciação  $d$  e a ordem de médias móveis  $q$ . Um gráfico de tempo dos dados geralmente sugere se alguma diferenciação seja necessária. Se a diferenciação for solicitada, fazendo a diferença dos dados uma vez  $d = 1$ , e inspecione o gráfico de tempo de  $\nabla X_t$ . Se a diferenciação adicional for necessária, tente novamente a diferenciação e inspecione um gráfico de tempo de  $\nabla^2 X_t$ .

Além dos gráficos de tempo, o ACF amostral pode ajudar a indicar se a diferenciação é necessária. Como o polinômio  $\phi(z)(1-z)^d$  tem uma raiz unitária, o ACF amostral  $\hat{\rho}(h)$ , não decairá a zero rápido quando  $h$  aumenta. Assim, um decaimento lento em  $\hat{\rho}(h)$  é uma indicação de que a diferenciação pode ser necessária.

Quando os valores preliminares de  $d$  foram estabelecidos, o próximo passo é olhar para o ACF e o PACF amostrais de  $\nabla^d X_t$  para quaisquer valores de  $d$  que tenham sido escolhidos. Como estamos lidando com estimativas, nem sempre ficará claro se o ACF ou o PACF amostral está diminuindo ou cortando. Além disso, dois modelos aparentemente diferentes podem ser muito semelhantes. Com isso em mente, não devemos nos preocupar em ser tão precisos neste estágio do ajuste do modelo. Neste ponto, alguns valores preliminares de  $p$ ,  $d$  e  $q$  devem estar à mão, e podemos começar a estimar os parâmetros.

### Exemplo III.39. Análise de dados do PIB.

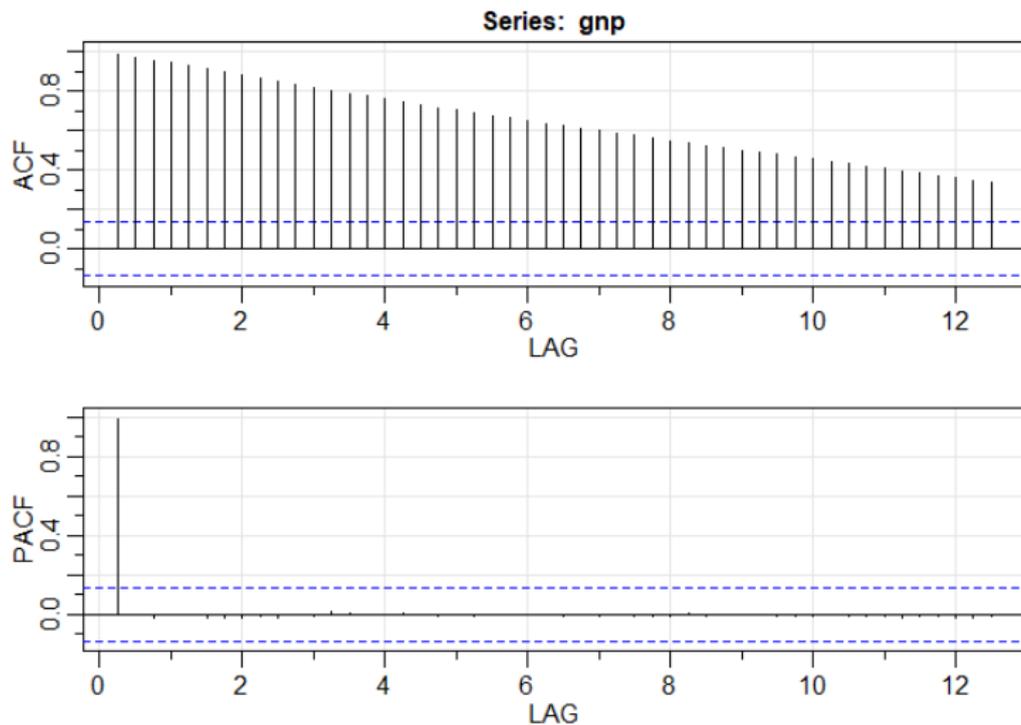
Neste exemplo, consideramos a análise do PIB trimestral dos EUA do primeiro trimestre de 1947 ao terceiro trimestre de 2002,  $n = 223$  observações. Os dados são o Produto Interno Bruto real dos EUA em bilhões de dólares de 1996 encadeados e foram ajustados sazonalmente. Os dados foram obtidos do Federal Reserve Bank of St. Louis. A figura abaixo mostra um gráfico dos dados, digamos,  $Y_t$ . Como a tendência forte tende a obscurecer outros efeitos é difícil ver qualquer outra variabilidade nos dados, exceto por grandes quedas periódicas na economia.



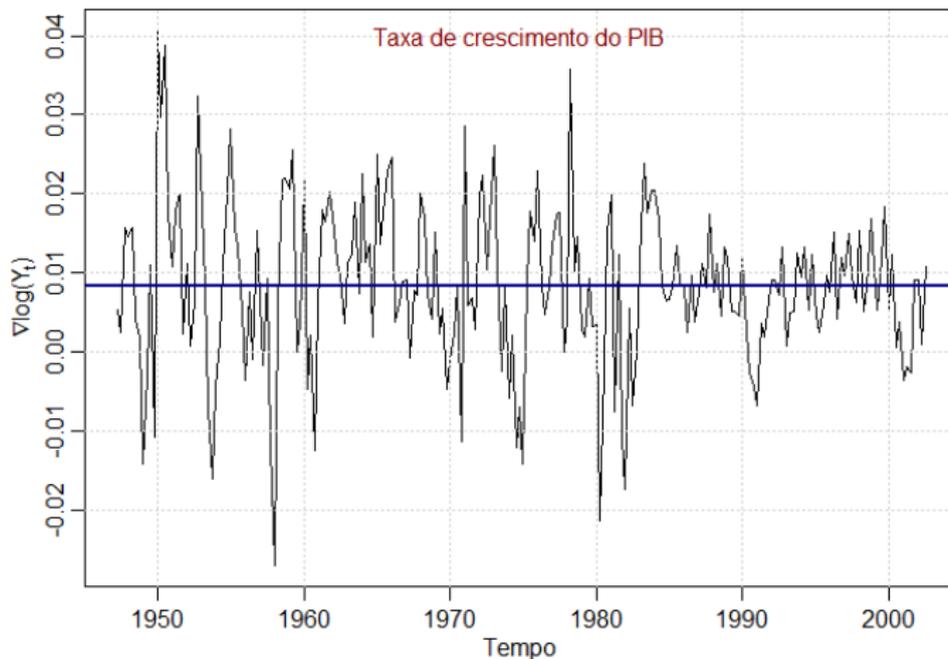
O código R para construirmos a figura acima e o ACF e PACF é o seguinte:

```
> par(mfrow = c(1,1), mar=c(3,3,1,1), mgp=c(1.6,.6,0), pch=19)
> plot(gnp, xlab="Tempo", ylab="Bilhões de Dólares",
      main="PIB trimestral dos Estados Unidos")
> grid()
> acf2(gnp, 50)
```

Quando são apresentados relatórios do PIB e indicadores econômicos similares, geralmente é na taxa de crescimento ou variação percentual e não nos valores reais ou ajustados que presta-se interesse. A taxa de crescimento, digamos,  $X_t = \nabla \log(Y_t)$  é plotada na figura embaixo e parece ser um processo estável.

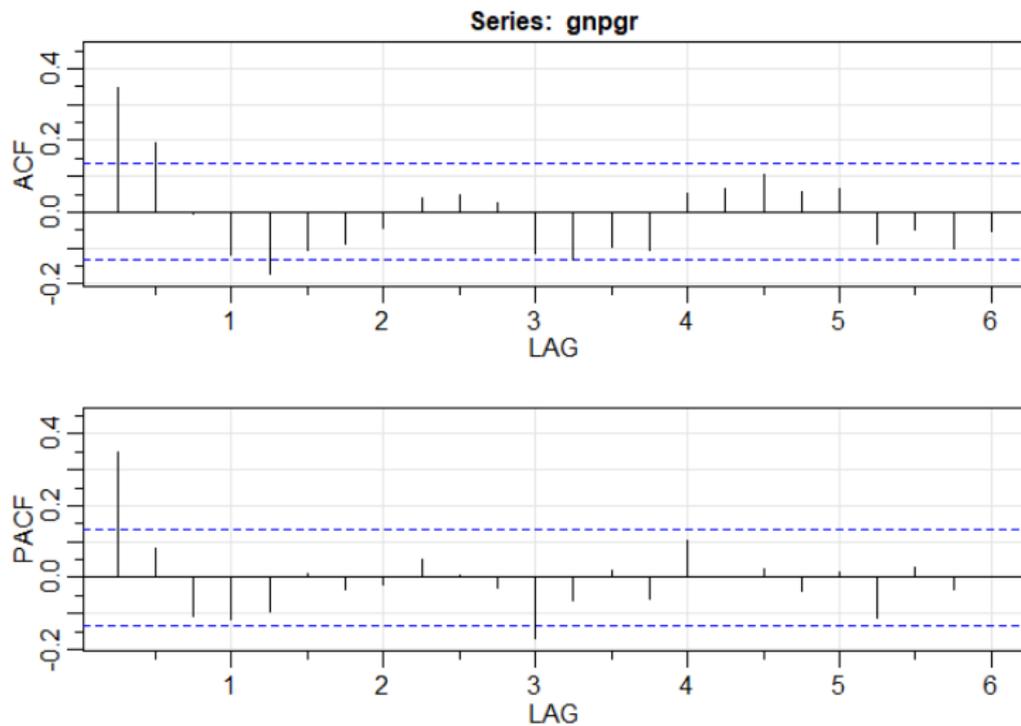


```
> gnpgr = diff(log(gnp)) # taxa de crescimento
> plot(gnpgr, xlab="Tempo", ylab=expression(paste(nabla,log,"(",Y[t],")
> abline(h=mean(gnpgr), lwd=2, col="darkblue")
> text(1975, 0.04, "Taxa de crescimento do PIB ", col="darkred")
```



A funções ACF e PACF amostrais da taxa de crescimento trimestral está representadas na figura abaixo. Inspeccionando as ACF e PACFa mostrais, podemos sentir que o ACF está cortando no lag ou atraso 2 e o PACF está diminuindo.

Isto sugeriria que a taxa de crescimento do PIB segue um processo MA(2) ou que o log do PIB segue um modelo ARIMA(0,1,2). Em vez de focar em um modelo, também sugeriremos que parece que o ACF está diminuindo e o PACF está cortando na defasagem 1. Isso sugere um modelo AR(1) para a taxa de crescimento ou ARIMA(1,1,0) para logartimo do PIB. Como uma análise preliminar, vamos encaixar ambos os modelos.



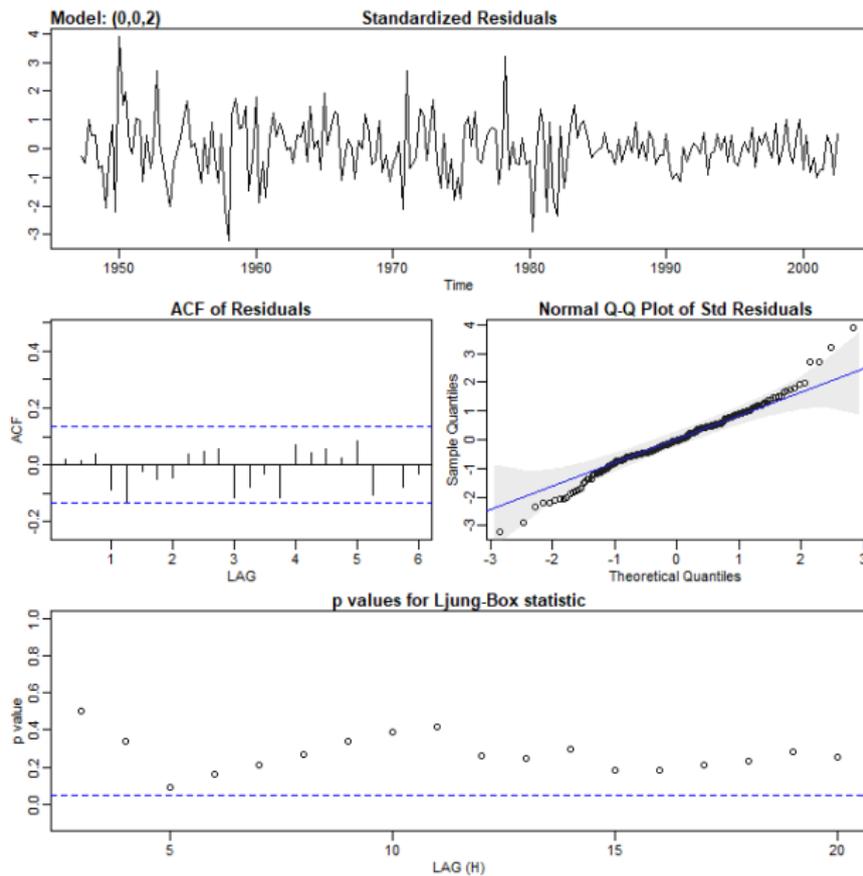
```
> acf2(gnpgr, 24)
```

Usando a estimação por máxima verossimilhança para ajustar o modelo  $MA(2)$  para a taxa de crescimento  $X_t$ , o modelo estimado é

$$\hat{X}_t = 0.008_{(0.001)} + 0.303_{(0.065)}\hat{W}_{t-1} + 0.204_{(0.064)}\hat{W}_{t-2} + \hat{W}_t,$$

onde  $\hat{\sigma}_w = 0.0094$  com 219 graus de liberdade. Os valores entre parênteses correspondem aos erros padrão estimados. Todos os coeficientes de regressão são significativos, incluindo a constante.

```
> sarima(gnpgr, 0, 0, 2) # MA(2)
```

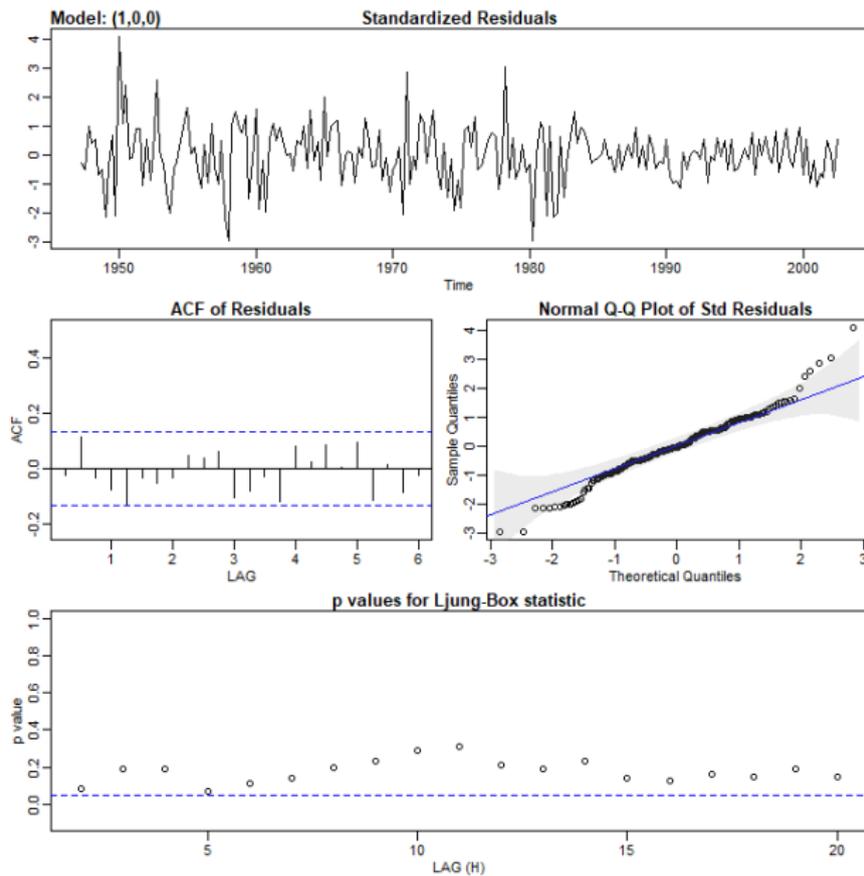


### Exemplo III.39. Análise de dados do PIB (continuação).

Fazemos uma nota especial disso porque, como padrão, alguns pacotes computacionais não ajustam uma constante em um modelo diferenciado. Ou seja, esses pacotes assumem, por padrão, que não há desvio. Neste exemplo, não incluir uma constante leva a conclusões erradas sobre a natureza da economia dos EUA. A não inclusão de uma constante assume que a taxa de crescimento média trimestral é zero, enquanto a taxa de crescimento trimestral média do PIB dos EUA é de cerca de 1%, o que pode ser visto facilmente na figura da taxa de crescimento. O modelo  $AR(1)$  estimado é

$$\hat{X}_t = 0.008_{(0.001)}(1 - 0.347) + 0.347_{(0.063)}\hat{X}_{t-1} + \hat{W}_t,$$

onde  $\hat{\sigma}_w = 0.0095$  com 220 graus de liberdade, observe que a constante na expressão acima é  $0.008(1-0.347)=0.005$ .



O próximo passo no ajuste do modelo é o diagnóstico. Esta investigação inclui a análise dos resíduos, bem como comparações de modelos. Mais uma vez, o primeiro passo envolve um gráfico de tempo das inovações ou resíduos  $X_t - \hat{X}_t^{t-1}$  ou das inovações padronizadas

$$\epsilon_t = \frac{X_t - \hat{X}_t^{t-1}}{\sqrt{\hat{P}_t^{t-1}}},$$

onde  $\hat{X}_t^{t-1}$  é a previsão de  $X_t$  com um passo à frente com base no modelo ajustado e  $\hat{P}_t^{t-1}$  é a variância estimada do erro de um passo à frente. Se o modelo se encaixa bem, os resíduos padronizados devem se comportar como uma sequência independente e identicamente distribuída com média zero e variância um. O gráfico de tempo deve ser inspecionado para quaisquer desvios óbvios desta suposição.

A investigação da normalidade marginal pode ser realizada visualmente, observando-se um histograma dos resíduos. Além disso, um gráfico de probabilidade normal ou um gráfico Q-Q pode ajudar a identificar desvios da normalidade.

Existem vários testes de aleatoriedade, por exemplo, o teste de execução, que pode ser aplicado aos resíduos. Poderíamos também inspecionar as autocorrelações amostrais dos resíduos, digamos,  $\hat{\rho}_\epsilon(h)$ , para quaisquer padrões ou valores grandes.

Lembre-se de que, para uma sequência de ruído branco, as autocorrelações amostrais são aproximadamente independentes e normalmente distribuídas com média zero e variância  $1/n$ . Portanto, uma boa verificação na estrutura de correlação dos resíduos é traçar  $\hat{\rho}_\epsilon(h)$  versus  $h$  juntamente com os limites de erro de  $\pm 2/\sqrt{n}$ .

Os resíduos de um ajuste de modelo, não terão exatamente as propriedades de uma sequência de ruído branco e a variância de  $\hat{\rho}_\epsilon(h)$  pode ser muito menor que  $1/n$ . Esta parte dos diagnósticos pode ser vista como uma inspeção visual de  $\hat{\rho}_\epsilon(h)$  com a principal preocupação sendo a detecção de desvios óbvios da suposição de independência.

Além de mostrar o  $\hat{\rho}_\epsilon(h)$ , podemos realizar um teste geral que leva em consideração as magnitudes de  $\hat{\rho}_\epsilon(h)$  como um grupo. Por exemplo, pode ser o caso de, individualmente, cada  $\hat{\rho}_\epsilon(h)$  ser pequeno em magnitude, digamos, cada um é ligeiramente menor que  $2/\sqrt{n}$  em magnitude, mas, coletivamente, os valores são grandes. A estatística Q de Ljung-Box-Pierce dada por

$$Q = n(n+2) \sum_{h=1}^H \frac{\hat{\rho}_\epsilon(h)^2}{n-h},$$

pode ser usada para realizar tal teste.

O valor  $H$  na expressão acima é escolhido arbitrariamente, tipicamente,  $H = 20$ . Sob a hipótese nula de adequação do modelo, assintoticamente,  $Q \sim \chi_{H-p-q}^2$ . Assim, rejeitaríamos a hipótese nula no nível  $\alpha$  se o valor de  $Q$  exceder o valor do quantil  $1 - \alpha$  da distribuição  $\chi_{H-p-q}^2$ . A ideia básica é que se  $W_t$  é ruído branco, então por Proposição I.2,  $n\hat{\rho}_W^2(h)$  para  $h = 1, \dots, H$ , são variáveis aleatórias assintoticamente independentes com distribuição  $\chi_1^2$ . Isso significa que

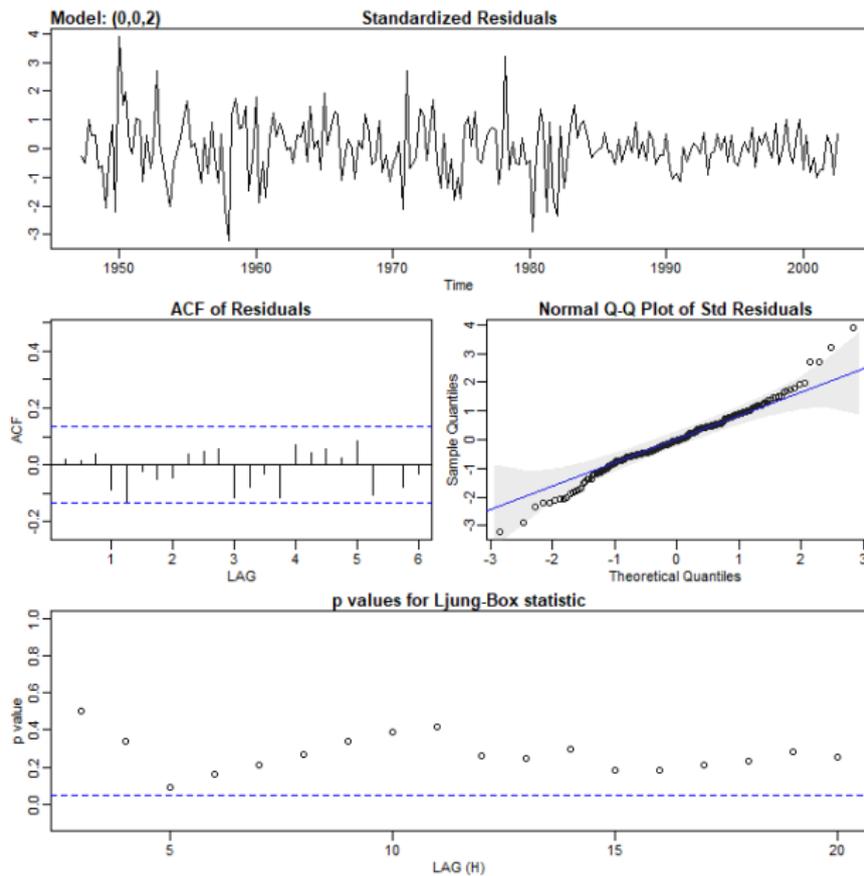
$$n \sum_{h=1}^H \hat{\rho}_W^2(h),$$

é aproximadamente uma variável aleatória com distribuição  $\chi_H^2$ . Como o teste envolve o ACF dos resíduos de um ajuste de modelo, há uma perda de  $p + q$  graus de liberdade; os outros termos são usados para ajustar a estatística para melhor corresponder à distribuição qui-quadrada assintoticamente.

### Exemplo III.40. Diagnósticos para o exemplo da taxa de crescimento do PIB.

Vamos nos concentrar no ajuste MA(2) do Exemplo III.39; a análise dos resíduos do modelo AR(1) é semelhante. A figura mostra um gráfico dos resíduos padronizados, o ACF dos resíduos, um boxplot dos resíduos padronizados e os p-valores associados à estatística-Q nas defasagens  $H = 3$  até  $H = 20$ , com  $H - 2$  graus de liberdade.

A inspeção do gráfico de tempo dos resíduos padronizados na figura não mostra padrões óbvios. Observe que pode haver outliers, com alguns valores excedendo 3 desvios padrão em magnitude. O ACF dos resíduos padronizados não mostra nenhum desvio aparente das premissas do modelo e a estatística Q nunca é significativa nas defasagens mostradas. O gráfico Q-Q normal dos resíduos mostra que a suposição de normalidade é razoável, com exceção dos possíveis outliers.



A etapa final do ajuste do modelo é a escolha do modelo ou a seleção do modelo. Ou seja, devemos decidir qual modelo iremos reter para previsão. As técnicas mais populares, AIC, AICc e BIC, foram descritas na Seção II.1 no contexto de modelos de regressão.

### Exemplo III.43. Escolha do modelo para a série do PIB dos EUA.

Retornando à análise dos dados do PIB dos EUA apresentados no Exemplo III.39 e Exemplo III.40, lembremos que dois modelos, um AR(1) e um MA(2), se ajustam bem à taxa de crescimento do PIB. Para escolher o modelo final, comparamos o AIC, o AICc e o BIC para ambos os modelos. Esses valores são um subproduto das execuções do comando `sarima` exibidas no final do Exemplo III.39 mas, por conveniência, exibimos novamente aqui lembrando que os dados da taxa de crescimento estão no objeto `gnpgr`:

```
> # AR(1)
> sarima(gnpgr, 1, 0, 0, details = FALSE)$AIC
[1] -6.44694
> sarima(gnpgr, 1, 0, 0, details = FALSE)$AICc
[1] -6.446693
> sarima(gnpgr, 1, 0, 0, details = FALSE)$BIC
[1] -6.400958
> # MA(2)
> sarima(gnpgr, 0, 0, 2, details = FALSE)$AIC
[1] -6.450133
> sarima(gnpgr, 0, 0, 2, details = FALSE)$AICc
[1] -6.449637
> sarima(gnpgr, 0, 0, 2, details = FALSE)$BIC
[1] -6.388823
```

AIC e AICc preferem o ajuste MA(2), enquanto o BIC prefere o modelo AR(1) mais simples. Geralmente, o BIC selecionará um modelo de ordem menor que o AIC ou AICc. Não é irracional manter o AR(1) porque os modelos autorregressivos puros são mais fáceis de trabalhar.